



# nature

## MOLECULAR MACHINES

Metallic boost  
for plasmas

PROTEIN  
ENGINEERING  
Keep it simple

WATER WARS  
Just a myth?

## FLORAL ATTRACTION

Electric pollen-tube  
attractant identified

NATURE.COM  
Focus on chemistry

# Turkey censors evolution

Turkey's government has done more for science than many. A row over a censored magazine and a sacked editor could put the good work at risk.

It has been the biggest crisis in Turkish academia since last year's lifting of the headscarf ban in universities. Last week a portrait of Charles Darwin was taken off the cover of the March issue of the government-backed science magazine *Bilim ve Teknik* (*Science and Technology*) just before it went to press. TÜBİTAK, Turkey's national science funding agency, which publishes the magazine, then sacked its editor, Çiğdem Atakuman. Scientists, assuming censorship, are justifiably outraged and protests are ongoing.

Science minister Mehmet Aydın, a historian of philosophy and religion, expressed discomfort at the cover's removal — but also fanned the flames by commenting: "What kind of a fight can we have with Darwin? The guy is already dead." He made matters worse by later adding: "[TÜBİTAK] is supposed to reflect the views of all those who have served science, no matter how mistaken they can be."

TÜBİTAK vice-president Ömer Cebeci, who sits on the magazine's editorial board, pulled the plug on Darwin. He denied censorship, charging that Atakuman had secretly changed an issue intended to cover global warming. Not true, says Atakuman, who says Cebeci told her that the Darwin cover was a "provocation" at a time of imminent local elections. One editorial-board member of *Bilim ve Teknik* has resigned in protest at what he, at least, considers censorship.

This row has brought into focus two issues that plague Turkish science. One is political interference in the scientific civil service; the other is high levels of public support for creationism.

In Turkey, as in many countries, the civil service is expected to mirror the ruling party's ideology. So, although they are keen funders of research, most senior government officials, in common with most of the population, do not believe in evolution by natural selection. The education minister Hüseyin Çelik, for example, has proclaimed his belief in intelligent design. Yet Turkey is one of three current

candidates for membership of the European Union (EU). Ankara and Istanbul house the largest overseas missions of the European Commission, whose officials are monitoring all aspects of Turkish public life and constantly advising on what needs to be done to harmonize laws and practices with those in the EU.

Science and technology was one of the first 'chapters' that the EU said it was satisfied with and that, provisionally, did not require further reform. Officials will now almost certainly be alarmed to see the extent of political interference in TÜBİTAK.

Only last week, the European Parliament issued a report stating that Turkey needed to make much faster progress in areas such as censorship. European officials will see this latest episode as evidence that the country has some way to go on that score.

TÜBİTAK needs to initiate a transparent investigation into the *Bilim ve Teknik* affair. The organization should also consider making an unambiguous statement of its position on evolution, intelligent design and creationism to reconfirm its credentials as a serious scientific body. In the past, TÜBİTAK has provided reliable information on Darwin's theory in a country where creation is offered as an alternative to evolution in high-school biology teaching. The agency could do that again. After all, none of the world's religions commands its believers to be creationists. Many Islamic scholars and thinkers have speculated on the origins of life.

Turkey's ruling party must learn from this latest affair. It must keep religion out of science policy, and be seen to be doing so. ■

**"Although they are keen funders of research, most senior officials, in common with most of the population, do not believe in evolution."**

## Health highway

President Obama's funds for electronic health records should prompt research — and controversy.

One of the most striking, if little noted, aspects of the recent US stimulus bill is that 53 of its 407 pages are devoted to setting up the administrative framework for a national system of electronic health records (EHRs). At the head of this framework, which encompasses committees on standards, policy and much else, will be a National Coordinator for Health Information Technology, located within the Department of Health and Human Services.

Many will find it even more remarkable that this portion of the stimulus bill builds upon foundations laid by former president George W. Bush, who established the office of the national coordinator by

executive order in 2004, and set him the goal of giving every American access to EHRs by 2014. Now, with the stimulus bill, Congress and the Obama administration have not only embraced that goal, but have allocated \$19 billion to help reach it.

Better still, from the research perspective, the stimulus bill explicitly envisions scientists being able to use these clinical records for a wide variety of population studies — monitoring the spread of flu, tuberculosis and other diseases, for example, or spotting people who might be good candidates for clinical trials, or looking for signs of adverse reactions after a drug has been released on the market (see page 278).

Many researchers believe that such studies could ultimately produce benefits as least as big as the effects of electronic records on clinical care — but only if some major issues about data access get resolved first.

Imagine, for example, that Mrs Smith checks into the hospital for hip-replacement surgery. Her doctors will be able to call up the electronic records of her personal physician, her physical therapist, her pharmacist and everyone else who provides her with health care. But this scenario poses some knotty questions. Even leaving aside the technical challenge of securing all those far-flung databases against attacks by hackers, who has the right to access any given database? And how much are they allowed to see there — everything, or just those portions of the data they need for their jobs, keeping the rest private? And who decides? Patients? Doctors? Hospitals? Anonymization and pseudonymization, in which identities are masked but all data can still be ascribed to respective individuals, protect individuals while enabling research, but how should such measures best be implemented?

These questions have been contentious even in the United Kingdom, which has deployed an electronic records system through its unified National Health Service. They will be even more contentious in the United States, where the health system is anything but unified, and where privacy concerns run deep.

The scientific questions are just as knotty. For example, will each patient have to give permission each time a researcher wants to access his or her records? That sounds reasonable — and indeed, this is the

spirit of the strict data privacy and security provisions of the federal Health Insurance Portability and Accountability Act. But that would make using the records for, say, routine epidemiology prohibitively cumbersome. There needs to be a clear policy allowing for such uses, as long as the researchers have no access to any aspect of the data that reveal a person's identity. But at the same time, the health-records system should provide individuals with an audit trail that shows them who has been looking at their records, and what they have seen there.

Or consider another scenario: a research group that wants to do a study aggregating EHR data over a large area. Will they have to get approval from every Institutional Review Board (IRB) — which oversee all research on human subjects in the United States — in every hospital and university involved? Common sense suggests that one IRB process is enough — and then only if the research uses personally identifiable data. But mechanisms to coordinate the oversight will have to be created.

The national coordinator and his advisory committees will have their hands full with these and many other issues. But the good news is that a management framework is in now place to address them, and to build an EHR system that will have the confidence of professionals and public alike. ■

## Filling the void

As science journalism declines, scientists must rise up and reach out.

Scientists at CERN, Europe's particle-physics laboratory near Geneva, Switzerland, opened the wine last week to celebrate the twentieth anniversary of the laboratory's invention of the World Wide Web. The scientists were also joined by around 60 members of the media, who may have been in a less festive mood. Even before the current economic crisis, the web was inflicting much pain on the mass media. Circulations have dropped, advertising has dried up and newspapers have been forced to lay off reporters and scale back coverage. A similar slump has hit the broadcast market, with no end in sight.

Science journalism is one of the numerous casualties in this media meltdown. Many science journalists are losing their jobs, and those who remain are being asked to provide content for blogs, podcasts, online videos and other new media (see page 274). Although it is difficult to know what effect these cutbacks have had on the public's understanding of science, the general feeling is that the quality of science coverage in the conventional media is declining — as is the media's ability to play a watchdog role in science, ferreting out fraud or other misconduct.

True, there is no shortage of scientific information on the web. Witness the way that research funding agencies use the web to inform the public about everything from planetary missions to public health. In principle, anyone with an Internet connection now has access to more, and better, scientific coverage than ever before.

In practice, however, this sort of information reaches only those

who seek it out. An average citizen is unlikely to search the web for the Higgs boson or the proteasome if he or she doesn't hear about it first on, say, a cable news channel. And as mass media sheds its scientific expertise, science's mass-market presence will become harder to maintain.

Harder, but not impossible. For example, scientists are blogging in ever increasing numbers, and the most popular blogs draw hundreds of thousands of readers each month. These blogging scientists not only offer expertise for free, but have emerged as an important resource for reporters. A *Nature* survey of nearly 500 science journalists shows that most have used a scientist's blog in developing story ideas. And a handful of universities, meanwhile, have started environmental publications that are run jointly by scientists and journalists. These publications seek to provide their journalistically valid, scientifically accurate content free of charge to the mainstream press.

Sadly, these activities live on the fringe of the scientific enterprise. Blogging will not help, and could even hurt, a young researcher's chances of tenure. Many of their elders still look down on colleagues who blog, believing that research should be communicated only through conventional channels such as peer-review and publication. Indeed, many researchers are hesitant even to speak to the popular press, for fear of having their carefully chosen words twisted beyond recognition.

But in today's overstressed media market, scientists must change these attitudes if they want to stay in the public eye. They must recognize the contributions of bloggers and others, and they should encourage any and all experiments that could help science better penetrate the news cycle. Even if they are reluctant to talk to the press themselves, they should encourage colleagues who do so responsibly. Scientists are poised to reach more people than ever, but only if they can embrace the very technology that they have developed. ■

# RESEARCH HIGHLIGHTS

## Fiery fountains

*Geology* **37**, 219–222 (2009)

On Earth, volcanic 'fire fountains' burst from the ground, driven by the fizz of carbonated magma rushing up narrow-necked vents. The same spectacular displays occurred on the Moon 3 billion–4 billion years ago, despite big differences in gravity, magma composition and magma viscosity, say Malcolm Rutherford of Brown University in Providence, Rhode Island, and Paolo Papale of the National Institute of Geophysics and Volcanology in Pisa, Italy.

By studying data collected from lunar samples, the researchers found that gases — primarily carbon monoxide rather than the carbon dioxide of terrestrial magma — form by graphite oxidation in lunar magma at a depth of about 8 kilometres. Modelling the upwelling for lunar conditions, they found that the magma escapes at the surface at about the same speeds as on Earth, and with similar gas volumes.



CORBIS

## EVOLUTION

### Painful pairing

*Curr. Biol.* **19**, 404–407 (2009)

Evolution of male sexual traits that harm females can be due to some reproductive benefit afforded the male by the harm, or to a negative side effect of a gene with some other positive benefit.

Cosima Hotzy and Göran Arnqvist of Uppsala University in Sweden examined the relationship between the vicious spines on the genitalia of male seed beetles (*Callosobruchus maculatus*), the damage these cause to females during mating, and sperm competition success.

They found a positive correlation between spine size and both harm to females and male fertilization success. However, statistical modelling showed that spine size still correlated with sperm success when harm was kept constant. This suggests that harm to females is simply an unfortunate side effect of the spines.

## CHEMISTRY

### Buckyball expansion

*Angew. Chem. Int. Edn* doi:10.1002/anie.200805870 (2009)

'Buckyball' molecular structures aren't confined to carbon. Uranium fullerenes that contain various numbers of uranium atoms have now joined the club, thanks to Peter Burns and his colleagues at the University of Notre Dame in Indiana.

Fullerene U60 (pictured right, top) looks just like its carbon analogue, whereas U44 (bottom) is peanut-shaped, comprising 12 hexagons and 12 pentagons. Fullerenes usually have as few adjacent pentagons as possible,

but the peanut bucks this trend. This is because the peanut shape has higher symmetry, which the uranyl clusters prefer over minimizing pentagon neighbours, the authors say.

Uranium fullerenes are stable in solution for several months and easy to make by a self-assembly process in water — in contrast to carbon fullerenes, which have to be zapped into being with a huge current.

## CANCER THERAPEUTICS

### Socking it to melanoma

*J. Clin. Invest.* doi:10.1172/JCI34015 (2009)

Human melanoma is resistant to treatment with retinoic acid, which is effective against several other cancers. But by activating the transcription factor SOX9, researchers at the National Cancer Institute in Bethesda, Maryland, may have discovered a way to make retinoic acid effective after all.

In a series of cell culture and tissue experiments, Vincent Hearing and his

colleagues identified mechanisms by which promoting SOX9 activity slows cell division and sensitizes melanoma cells to the drug. In mice injected with melanoma cells, a drug that activates SOX9 had mild effects on its own, but when combined with retinoic acid significantly reduced the size of tumours.

## GENOMICS

### Staying in shape

*Science* 10.1126/science.1169050 (2009)

Researchers have wondered why some regions of DNA that do not code for proteins but nevertheless are functionally important have sequences that have not been maintained by evolution. It could be because evolution is instead working to conserve the DNA's structure, say Elliott Margulies at the National Human Genome Research Institute in Bethesda, Maryland, Thomas Tullius at Boston University in Massachusetts and their colleagues.

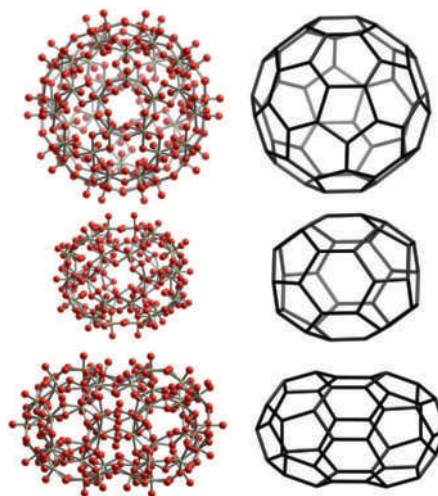
A DNA helix can have subtle structural variations in its grooves and its physical arrangement that affect how well proteins bind to it. By comparing a portion of DNA structure across 36 different species, the team estimates that 12% of the human genome is conserved — twice the estimate based on sequence conservation. The method should help to identify new non-coding regions with important biological roles.

## SEXUAL SELECTION

### Cheap flights

*Proc. R. Soc. B* doi:10.1098/rspb.2009.0090 (2009)

A flashy tail is often a way for a male bird to attract a mate, but the adornment has been assumed to make flying more arduous.





To find out whether this is true, Christopher Clark and Robert Dudley at the University of California, Berkeley, trained Anna's hummingbirds (*Calypte anna*) to fly in a wind tunnel with and without hugely elongated feathers from another species tacked onto their tail feathers. Elongated tails slowed the birds' average top speed from 15.1 to 14.6 metres per second. The increased metabolic cost of flight at high speeds was about 10%, but negligible at lower speeds.

The researchers suggest that by hiding tails aerodynamically in the wake of the body, diverse sexual signals can be kept relatively cost-free.

## GENOMICS

### Big is beautiful

*PLoS ONE* 4, e4688 (2009)

Finding gene regulators in the fruitfly *Drosophila melanogaster* has proved difficult: looking for conserved non-coding DNA sequences, a method that works well in vertebrates, has been unsuccessful. Michael Eisen at the University of California, Berkeley, and his co-workers suggest that this is because the fly's genome is compact — not because its regulatory architecture differs substantially from that of vertebrates.

Eisen's group sequenced various loci from four species of tephritid — 'true' fruitflies — which have large genomes containing islands of conserved non-coding DNA sandwiched between unconserved stretches. This allowed the researchers to pinpoint six conserved tephritid non-coding regions that functioned as gene enhancers in *D. melanogaster* embryos. Until now, the compact size of the *D. melanogaster* genome has been considered a boon to scientists.

## MOLECULAR IMAGING

### Nailing the molecule

*Science* 323, 1464–1468 (2009)

Molecules in gases and liquids are always moving thanks to their thermal energy. This means that measuring many details of their dynamics, which would require accessing the interaction of the molecule with nearby molecules and light, is impossible. However, by using a short laser pulse, a molecule can be 'frozen' for a few picoseconds ( $10^{-12}$  seconds), a time sufficient to examine, using ultrafast spectroscopy, the characteristics of its molecular dynamics.

Albert Stolow of the Steacie Institute for Molecular Sciences in Ottawa, Canada, and his colleagues demonstrate their method with the carbon disulphide molecule, observing its dynamics in a photochemical reaction.

## EVOLUTION

### Fishy fangs

*Proc. R. Soc. B* doi:10.1098/rspb.2009.0141 (2009)

Carp-like cypriniform fishes lost their teeth millions of years ago, but a newly described species has come close to redeveloping them.

Males of the tiny *Danionella dracula* have a series of fang-like pieces of bone projecting from their jaws, report Ralf Britz of London's Natural History Museum and his colleagues. So far, the maximum size reported for the fish, from Myanmar, is 16.7 millimetres. Its development also appears stunted; it retains many larval body features into adulthood and lacks 44 bones or parts of bones seen in its close relative the zebrafish (*Danio rerio*).

Remaining mysteries include the function of the toothy projections and whether they are formed through the genetic pathway for true tooth formation, which is known to be conserved in zebrafish.



R. BRITZ

## MATERIALS

### Five-star ice

*Nature Mater.* doi:10.1038/nmat2403 (2009)

As a snowflake's sixfold symmetry shows, ice crystals contain water molecules locked into hexagonal arrays.

But Angelos Michaelides at the London Centre for Nanotechnology and his colleagues have identified circumstances in which ice breaks this rule. Water molecules forming nanometre-wide chains on a particular copper surface at 100–140 kelvin are arranged in pentagons. The motif arises from a balance of hydrogen-bonding interactions between water molecules and the strain induced by fitting to the lattice of the copper surface beneath.

#### Correction

The Research Highlight 'Pogo-stick pictures' (*Nature* 458, 11; 2009) described the subject of the accompanying image as a mouse cochlear hair cell. It is in fact a hippocampal neuron.

## JOURNAL CLUB

Ahmad M. Khalil  
Harvard Medical School

**A geneticist views two theories of X-chromosome inactivation in a broad context.**

Female mammals have two X chromosomes, one of which is inactivated to ensure that females get the same dose of X-linked genes as males. Two long, non-coding RNA molecules mediate this process. One, *Xist*, initiates silencing of the chromosome to be inactivated, whereas its antisense partner, *Tsix*, blocks such silencing of the remaining X. The exact mechanism by which *Xist* and *Tsix* exert their functions is not known.

RNA interference (RNAi) describes a process by which RNA fragments 'interfere' with the creation of proteins from their RNA recipes. The recent discovery of RNAi in mammalian cells made it tempting to postulate that sense and antisense non-coding RNA partners, such as *Xist* and *Tsix*, are processed by the RNAi-associated enzyme DICER into small RNAs. Jeannie Lee and her team at Harvard Medical School have presented several lines of evidence in support of this, including data that demonstrate improper X inactivation in DICER's absence (Y. Ogawa *et al. Science* 320, 1336–1341; 2008).

By contrast, a study by David Livingston of Harvard's Dana-Farber Cancer Institute and his colleagues has shown that, in the absence of DICER, the inactive X chromosome in stem cells remains coated with *Xist* and several other repressive markers (C. Kanellopoulou *et al. Proc. Natl Acad. Sci. USA* 106, 1122–1127; 2009). This group suggests that the effects observed by Ogawa *et al.* may have been indirect because DICER is also involved in the processing of a class of small, non-coding RNAs known as microRNAs that can alter gene expression by fine-tuning protein production.

Despite their differences, these studies should provide an incentive to further investigate the potential role of the RNAi pathway in the nucleus of mammalian cells. This will shed light on both X inactivation and gene regulation in general.

Discuss this paper at <http://blogs.nature.com/nature/journalclub>

## NEWS

# Europe's green billions

Clean-energy investments target eastern Europe, but it remains to be seen whether they will create as many new jobs as hoped.

Three months after European Union (EU) leaders committed themselves to cutting greenhouse gases by at least 20% from 1990 levels by 2020, the European Commission has announced that it will subsidize 'green' programmes and infrastructures with €105 billion (US\$136 billion).

The windfall comes from the union's €347 billion in 'structural' funds. It is the second-largest budget in the commission's portfolio, behind agriculture, and is aimed at creating equal standards of living across the EU.

Although no fresh money is involved — the size of the structural funds was agreed in 2007 — the amount allocated to green projects is nearly triple that of the previous budget period, from 2000 to 2006. Regions within the EU are free to decide how to spend the money, as long as they meet the general criteria and priorities set by the commission.

The money is necessary if Europe is to deliver on its high hopes for wind and other low-carbon energy sources, says Greg Arrowsmith, a policy officer with the European Renewable Energy Centers Agency in Brussels. Just €2.3 billion of the EU's €51-billion Seventh Framework Programme — Europe's main

source of research funding — is earmarked for non-nuclear energy technologies.

The new investment will not cover carbon capture and storage projects, but it does address many others related to energy and climate. Around €9 billion is targeted for renewable energy and energy-efficiency projects, and an additional €29 billion will be invested in modernizing railways and in urban transport projects aimed at reducing emissions from cars. Significant funding will also be available for water and waste-management projects, biodiversity and nature conservation, and for encouraging the development of eco-friendly goods and services by small and medium-sized enterprises.

The commission hopes that the investment will create new jobs and help to revive regions that are suffering in the current economic downturn. "Support for the green economy and the environment goes hand-in-hand with the cohesion policy's objective to deliver sustainable growth, jobs and competitiveness," regional policy commissioner Danuta Hübner said when announcing the figures on 9 March.

The bulk of the money (€60 billion) will go to countries in eastern and southeastern

**"The massive public stimulus will give a lot of extra impetus to green investment."**



Europe that lag behind the rest of the union, economically and environmentally. The largest beneficiaries will be Poland (€18 billion), Spain (€12 billion), the Czech Republic (€11 billion) and Hungary (€9 billion). Money from the regional funds could, for example, be used to convert old ports on Poland's Baltic coast into transit points for testing and installing offshore wind turbines. It is, however, up to local politicians to decide exactly how to apportion the funds.

The plan exceeds the expectations of many economists and energy experts, who had hoped that the economic crisis would create an impetus for 'greening' the economy, as has happened with massive investments in countries

B. STRONG/REUTERS

## Society sues journal over right to reply

The Max Planck Society (MPS) in Germany has begun legal proceedings against publishers Wiley International in a dispute over an editorial in the February issue of *Human Brain Mapping*.

The society alleges that the editorial grievously misrepresents it and harms the reputation of one of its scientists. It wants the journal to publish a letter from the society addressing these concerns without delay.

Peter Fox, an editor-in-chief at *Human Brain Mapping*, says that the MPS letter went through normal

refereeing processes "in a timely manner", but says he does not know when it will be published. MPS vice-president Herbert Jäckle, a developmental biologist who was deputized to speak for the society, claims that the journal has unfairly delayed the society's right to reply.

The dispute has been raging for nearly a year. Fox, a neurologist at the University of Texas Health Science Center in San Antonio, accepted the letter on 11 March, two days after the deadline the society set before taking legal action. Fox says that he will publish the letter

together with a reply that "rebutts Dr Jäckle's various accusations".

"I don't have a problem with the journal printing a reply," says Jäckle. "But I fear that without expedited publication, the letter would appear on the web only after months of delay and in print only in a year."

The row originally centred on a dispute over who owned data gathered in the laboratory of Nikos Logothetis at the Max Planck Institute for Biological Cybernetics in Tübingen. Last spring, Amir Shmuel and David Leopold used the data in a paper for a special issue of *Human*



MAX PLANCK SOCIETY

Herbert Jäckle wants the Max Planck Society's letter published.

*Brain Mapping* on spontaneous brain activity — activity that occurs when there is no visual input. Logothetis





Wind farms could benefit from greater investment by the European Commission.

such as the United States, Japan, South Korea and Germany.

"This is the way to go," says Claudia Kemfert, an expert on climate, energy and transport at the German Institute for Economic Research in Berlin. "This kind of economic stimulus makes a lot more sense than bonuses to people who give up old cars to buy new vehicles." A €2,500 'car-scrapping' incentive, part of Germany's own €50-billion stimulus package, is boosting car sales, but has been criticized on economic and environmental grounds.

Oliver Schäfer, policy director at the Brussels-based European Renewable Energy Council, says that the EU's renewable-energy industry — involving some 350,000 jobs,

and generating €40 billion a year in business volume — will benefit substantially from the boost in public funding. Experts have estimated that it will take around €600 billion to help the EU increase the amount of energy it obtains from renewable sources from the current 8.5% to its goal of 20% in 2020. Subsidies through the structural funds programme will provide start-up financing for new solar, hydro- and wind-power facilities, for wires needed to transport 'green' electricity to consumers, and for decentralized energy supplies such as through rooftop solar-energy collectors.

But promises of funding injections have hit stumbling blocks in the past. The commission's economic recovery plan, proposed last November, foresees €3.5 billion in unspent money from the past budget period being reallocated to investment in energy technologies, including €500 million for offshore wind projects and €1.5 million for carbon capture and sequestration. But several groups in the European Parliament, which will vote on the issue later this month, are seeking to stop the plan. Critics say that few of the envisaged projects will materialize soon enough to help economic recovery.

Regardless, the €105 billion from the structural funds will still flow. "Sustainable funds have had renewable energies and water projects on their radar for quite some time," says Rolf Häßler, an analyst with the rating agency Oekom Research in Munich, Germany. "The massive public stimulus will give a lot of extra impetus to green investment." ■

Quirin Schiermeier

says that the data were used inappropriately because they had been gathered for a different purpose, when the research monkeys were staring at a flickering screen. He wrote to the journal's editor, among other actions, asking for the paper not to be published<sup>1</sup>.

It was, however, published online<sup>2</sup> on 8 May 2008. Logothetis published a scientific rebuttal in *NeuroImage*<sup>3</sup>, which appeared online in January 2009.

Fox, together with the three other editors of the special issue, published an extended editorial analysing the dispute in the framework

of ethics. This appeared online on 9 December 2008 and in print in the February issue of *Human Brain Mapping*<sup>4</sup>.

The MPS says that there are numerous factual errors in the editorial's account of events. Jägle is also incensed by the article's claims that he had given permission to publish the disputed paper. "I was only a mediator in a dispute, not an adjudicator as the editorial claims," he says. "It is not up to the Max Planck Society to permit publication of anything — that responsibility lies solely with the editors."

The MPS has posted a response to the February

editorial on the website of Logothetis's laboratory (see <http://tinyurl.com/c4yq9m>).

"The goal of our editorial was to use this conflict to discuss the ethical principles that govern responsible conduct of research and peer review," says Fox, "and thereby to develop guidelines that would prevent future occurrences of this nature." ■

Alison Abbott

1. Abbott, A. *Nature* **454**, 6-7 (2008).
2. Shmuel, A. & Leopold, D. A. *Hum. Brain Mapp.* **29**, 751-761 (2008).
3. Logothetis, N. K. et al. *NeuroImage* doi:10.1016/j.neuroimage.2009.01.010 (2009).
4. Fox, P. T. et al. *Hum. Brain Mapp.* **30**, 347-354 (2009).



#### CLINICAL TRIALS

European directive strangles research with red tape, scientists warn.  
[www.nature.com/news](http://www.nature.com/news)

PUNCHSTOCK

## Scientists in bone battle

Officials at the University of California are moving to give two of the oldest-known skeletons in North America to a local Native American tribe, against the recommendation of university scientists who say the bones should be retained for study.

Under federal law, bones are returned to a tribe that can prove 'cultural affiliation' through artefacts or other analyses. At nearly 10,000 years old, the skeletons in question — unearthed in 1976 at the University of California, San Diego (UCSD) — are so ancient that they are not culturally linked to any tribe (see *Nature* 455, 1156-1157; 2008).

But last month, University of California president Mark Yudof and UCSD chancellor Marye Anne Fox began seeking a rare federal approval to give the skeletons to the local Kumeyaay tribe, which has asked for them. And some anthropologists say the decision is based on politics, not science.

"This is scandalous," says Robert Bettinger, an anthropologist at the University of California, Davis, who is on the panel that oversees how archaeological remains are handled at all ten University of California campuses. The panel was not consulted on this transfer proposal.

"This goes against the policies of President Barack Obama for science-based decisions, not belief-based ones," says anthropologist Margaret Schoeninger, who chairs the UCSD committee that reviews such specimens. The final decision is likely to be made by Obama appointee Ken Salazar, secretary of the Department of the Interior.

Officials at the Department of the Interior say the request will be considered by the committee that reviews the return of remains and artefacts when it meets in Seattle, Washington, on 23-24 May.

Fox declined an interview, but said in a statement that the transfer "seems an appropriate balance between the interests of science and [those] of the Native American community".

This week, Schoeninger is petitioning the research committee of the UCSD faculty senate to try to prevent the transfer. But Kumeyaay spokesman Steve Banegas says the tribe impatiently awaits the chance to bury the bones. "This is long overdue," he says. ■

Rex Dalton

## SPECIAL REPORT

# America's long hard road to climate-change law

What challenges lie ahead as the United States tries to construct a working system for greenhouse-gas regulation? **Jeff Tollefson** reports.

"We're back." That was the message from Todd Stern, the United States' new international envoy on climate, to congressional staff, scientists and representatives of several countries on 3 March in Washington DC.

In a room overlooking the Capitol dome, Stern noted the United States' responsibility as the historical emitter of most greenhouse gases globally. He explicitly recognized the right of developing nations to lift their people out of poverty, and pledged resources to help them make the transition to clean development.

But Stern — the person charged with leading US negotiations through the United Nations climate-change talks — went on to challenge Europe on multiple fronts. Most notably, he suggested that Europe's proposal for industrialized nations to reduce their emissions by 25–40% below 1990 levels by 2020 is unfair, unnecessary and "a prescription for stalemate".

President Barack Obama has pushed for a cap-and-trade system to control greenhouse gases in which industries can buy and sell permits, auctioned off from the start, in order to stay within an overall cap on pollution. It is not yet clear whether the United States will embrace such a strict regime. But for Obama, the place to start is working with Congress.

Congressional leaders have talked before about legislation to create a US cap-and-trade system, but the bills usually never make it out of committee to be voted on by all lawmakers. This year, however, may be different. With the president on board and strong Democratic majorities in both congressional houses, 2009 may be the year when cap-and-trade legislation gets done. The question on Capitol Hill is not so much if but when Obama will sign such a bill into law — and how that law will mesh with US negotiations for the next international climate agreement. In December, international envoys will meet in Copenhagen to hammer out a successor to the Kyoto Protocol on Climate Change. Stern will represent Obama there, and he knows the necessity of working with Congress. In 1997, even as he was helping negotiate the final details of the Kyoto Protocol, the Senate signalled its discontent with a

symbolic 95–0 vote against the treaty.

Europe has recently been streaming emissaries through Washington DC in an effort to build alliances and provide perspectives on how to set up a cap-and-trade system. And they are actively engaging members of Congress and policy-makers instead of just focusing on Obama. "This is too important to be left to the president," says Ed Miliband, the UK Secretary of State for Energy and Climate Change. "The Congress needs to be an equal partner."

Obama has appointed an all-star cast to develop a federal strategy on climate and to work closely with Congress on the issue (see 'The key players'). No one is yet sure how well this assembled cast will work to advance climate policy. "But there are some very smart and very capable people who have gone into the administration, who have a deep knowledge of climate policy and who have all of the talent and capability to get this done," says Tim Profeta, who heads Duke University's Nicholas Institute for Environmental Policy Solutions in Durham, North Carolina.

The first moves from the administration might come as early as this month, when the Environmental Protection Agency could announce its long-delayed response to a 2007 Supreme Court decision establishing the agency's authority to regulate carbon dioxide as a pollutant. Lisa Jackson, the agency's administrator, is almost certain to issue an 'endangerment' finding declaring CO<sub>2</sub> a pollutant that should be regulated under the 1970 Clean Air Act. Technically, such a move would allow the Obama administration to move forward with regulating carbon on its own.

But Obama's main goal is to push a bill through Congress. The first actions from lawmakers could come from the House of Representatives. Henry Waxman (Democrat, California), chair of the House energy and commerce committee, has said his committee will produce a comprehensive energy and climate bill by the end of May. House speaker Nancy Pelosi (Democrat, California) has promised to bring such a bill out of committee and onto the full House floor for a vote this year; it would be

**"This is too important to be left to the president."**  
— Ed Miliband



the first-ever climate vote in the House.

Alden Meyer, who handles climate policy for the Union of Concerned Scientists in Washington DC, says that Waxman's approach of putting energy and climate into one bill makes sense. Like many, Meyer argues that issues such as energy efficiency, renewable energy and transportation are more easily addressed through separate regulations than through the cap-and-trade programme itself. "That's one of the things we've learned," says Meyer. "You need other policy mechanisms in addition to cap-and-trade."

It's not yet clear whether the Senate will take that same comprehensive approach. Last June, the Senate shot down a bill that would have reduced emissions to 63% below 2005 levels by 2050. Senate majority leader Harry Reid (Democrat, Nevada) has said he will bring a bill to the floor again this year, but there is still debate as to how comprehensive it might be — or not. Barbara Boxer (Democrat, California), who chairs the Senate's environment and public works committee, has said she wants her committee to produce "simplified legislation". That could well mean a cap-and-trade outline only, leaving other committees to handle provisions related to things such as energy, taxes and agriculture.

## Balancing act

Any energy issues would be headed by Jeff Bingaman (Democrat, New Mexico), who chairs the energy and natural resources committee in the Senate and who fears that progress on renewables might be held hostage by hammering out the more complicated details of cap-and-trade legislation. To overcome such blocks, Bingaman is working with a loose group of centrists from both parties, often dubbed the 'Gang of 16', members of which are searching for ways to make greenhouse-gas caps more politically palatable. Their staff





Congress is under increasing pressure to pass legislation regulating carbon dioxide emissions.

are looking at technical solutions that tackle the thorniest problems, including how to control costs, provide environmental integrity and ensure that domestic industries are not put at a disadvantage if, say, China fails to act on controlling its own emissions. "This whole debate often focuses on the differences between those who advocate clear environmental certainty and those who want clear economic certainty," says Jonathan Black, who handles climate for Bingaman on the energy committee.

The same debates have played out internationally, with some countries suddenly backedpedalling on commitments even as the United States steps forward. Canadian voters rejected a strong environmental platform proposed by the Liberal party last autumn, and cap-and-trade legislation is floundering in Australia despite support from Prime Minister Kevin Rudd. European leaders had to make concessions to various industries in December 2008 to hold together their coalition on the next phase of their carbon trading scheme, which begins in 2013.

If it had been successful, last June's climate legislation in the United States would have controlled costs by creating an independent federal board with the authority to intervene if the market gets out of control. Companies would also have been able to borrow some credits from the future, meaning they could emit more now, as long as they emit even less in subsequent years. Bingaman favours the use of set minimum and maximum prices for credits to ensure economic predictability, but such a 'price collar' would not guarantee emissions reductions because more credits could be issued if prices were to reach a set level.

**"The question is how much they are really willing to fight for this."**

Many lawmakers are also looking at border-tax adjustments — essentially tariffs — on imports if certain industries find themselves at an economic disadvantage as a result of carbon regulations. Another method for protecting specific industries is to allocate some emissions permits to them for free, which is how European regulators will handle the issue in the next phase of their emissions trading scheme.

The challenge of getting Congress to act becomes even more difficult if Stern actually brings home an international treaty from Copenhagen, which requires a two-thirds majority (67 votes) to pass in the Senate. "There are far less controversial agreements that have not managed to attract that level of supermajority here in the Senate," says Nigel Purvis, a former US climate negotiator who works on climate issues at the Brookings Institution in Washington DC.

That has Purvis, as well as some in Congress and the administration, thinking about 'congressional executive agreements'. Used for most international trade agreements, these have the same legal effect as a treaty, but require only 60 votes to pass in the Senate. In these, Congress would authorize the president to negotiate an agreement, which is then voted yes or no by Congress on its return.

Kathleen Frangione, chief climate aide to Senator John Kerry (Democrat, Massachusetts) on the Senate foreign relations committee, says that no matter what, the White House will have to be fully engaged if legislation is going to move this year. "The question is how much they are really willing to fight for this year," she says. "And we don't really know yet." ■

## THE KEY PLAYERS

Leaders in both houses of US Congress say they will bring climate legislation to a vote this year, so congressional staff are busy hammering out the details of such a bill — particularly a cap-and-trade solution to limit greenhouse-gas emissions. The Obama administration is pushing to get such legislation passed by lawmakers. The debate's key players are almost all Democrats, who expanded their majorities in Congress in the 2008 election.

### THE OBAMA ADMINISTRATION Carol Browner

Assistant to the president on energy and climate change

**Lisa Jackson**  
Administrator,  
Environmental Protection  
Agency

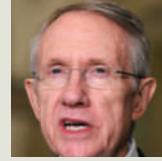
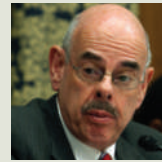
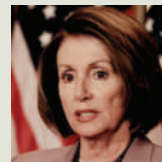
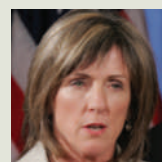
**HOUSE OF REPRESENTATIVES  
Nancy Pelosi**  
Democrat, California  
Speaker of the House

**Henry Waxman**  
Democrat, California  
Chairman of the House  
committee on energy  
and commerce

**SENATE  
Harry Reid**  
Democrat, Nevada  
Majority leader

**Barbara Boxer**  
Democrat, California  
Chairwoman of the  
Senate committee on  
environment and public  
works

**Jeff Bingaman**  
Democrat, New Mexico  
Chairman of the Senate  
committee on energy  
and natural resources



A. WONG/GETTY

A. POERTNER, ROSWELL DAILY RECORD/AP; ZUMA PRESS/NEWS.COM; K. DIETSCH/UPH PHOTO/NEWS.COM; C. KLEPONS/AFIP/GETTY; L. V. BURKE/AP PHOTO; AP

**NATUREJOBS**

US universities hit by recruitment freeze, p. 372.  
[www.naturejobs.com](http://www.naturejobs.com)

# Big interest in heavy drugs

The drug industry is seeking profits by modifying hydrogen in existing medications.

Pharmaceutical companies are beginning to bet on the idea that simply switching a hydrogen atom with a heavier isotope in a currently approved drug could create a better drug. Encouraged by results from clinical trials, companies are snapping up intellectual-property rights on many of the modified drugs.

On 16 March, for example, Concert Pharmaceuticals of Lexington, Massachusetts, reported results of a phase I clinical trial for a version of the antidepressant paroxetine, sold as Seroxat by GlaxoSmithKline and first marketed under that trade name in 1992. Concert's version swaps out one or more of the hydrogen atoms in the paroxetine structure (see 'Drug modifications') for deuterium, a heavier isotope of hydrogen that contains a proton and a neutron, rather than just a proton. The company was testing this deuterated version for treating hot flushes without the side effect of standard paroxetine in which the liver enzyme CYP2D6 is inactivated. Because this enzyme metabolizes many other drugs, inactivating it means it is hard to take other drugs along with paroxetine.

In Concert's trial, 94 women tested paroxetine — both standard and deuterated versions — along with the cough medicine dextromethorphan, which is metabolized by CYP2D6. Taken with standard paroxetine, metabolic uptake of the cough medicine is inhibited, says Roger Tung, chief executive at Concert. However, the deuterated version showed less metabolic inhibition — suggesting that the drug might be better to combine with others.

In theory, deuterated drugs can work differently in the body because deuterium can make stronger chemical bonds than hydrogen. This can affect how quickly a drug is broken down.

Another company working with deuteration



Substituting atoms in drugs could allow patenting opportunities and create better pharmaceuticals.

is Auspex Pharmaceuticals, based in Vista, California. Last October, Auspex announced phase I clinical trial results with a deuterated version of venlafaxine. Venlafaxine was first produced by Wyeth, in 1993, as the antidepressant Effexor. The trial suggested that the deuterated version of the drug stayed in the bloodstream longer than the non-deuterated version and may even cause fewer side effects, says Mike Grey, chief executive of Auspex.

The company already has a patent on its deuterated version of venlafaxine, and has filed more than 150 other US patents on families of deuterated drugs. Concert has filed more than 100 US patent applications and has received two notices of allowance already, and is expecting to get its first patents this year.

## Beyond the obvious

Kevin Mooney, a patent lawyer with Simmons & Simmons in London, says the strategy of piling up patent applications on deuterated versions of existing drugs is legitimate. "Everyone is entitled to research on other compounds," he says. Still, the onus will be on companies to demonstrate the usual criteria for patentability. "They would have to show that this deuterated form of paroxetine was new," he says, "then they would have to show that it wasn't an obvious thing to do." Patent applications — for example, those involving different salts of known compounds — often fail on this second requirement of non-obviousness.

Tung is confident that his company's tactics will work. "We treat this as an entirely new chemical entity," he says.

But in future, getting such patents will be harder, says Kirk Gallagher, a pharmaceutical patent lawyer at D Young and Co in London.

"The obviousness bar will be raised as time goes by," he says, as the idea of deuterating a drug to get different pharmacokinetic properties becomes commonplace. Rather than the broad patents covering families of drugs being applied for now, companies will have to show improved action of specific molecules to convince a patent-giver that they have something new, Gallagher says. "They'll have to do more science before they file their patents."

Derek Lowe, a drug-discovery scientist and author of the *In the Pipeline* blog, says larger drug companies may take the same approach. "I think that every big pharma company is keeping this in mind," he says. "That's where the start-ups are going to have a problem."

The approach with deuterated drugs is similar to that used by Sepracor, based in Marlborough, Massachusetts, in the 1980s. Until that point, patents didn't include specific information on isomeric forms of drugs. Sepracor made its business by filing patents on active isomers of known drugs. In response, pharmaceutical companies began to routinely specify chiral isomers on patents. "Pharma companies will now probably add deuterium analogues," says Gallagher.

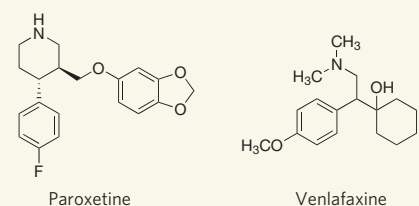
Grey isn't worried. "There is still a vast landscape for us to explore," he says. Auspex hopes to partner with larger drug companies to continue developing its products.

Concert is similarly looking for ways forward. The company has raised US\$96 million in start-up money so far, but has decided not to develop its deuterated paroxetine further for now — focusing instead on another deuterated drug, an HIV protease inhibitor, which it thinks will be a bigger seller.

**Katharine Sanderson**

## DRUG MODIFICATIONS

Pharmaceutical companies have replaced hydrogen atoms with deuterium to create versions of paroxetine and venlafaxine to test in clinical trials.





# Roche vows to keep Genentech culture

An era has ended for the biotechnology industry with the full merger of biotech pioneer Genentech, of South San Francisco, and Swiss pharmaceutical giant Roche, based in Basel.

The deal, announced on 12 March, ends Roche's eight-month pursuit of Genentech. Roche's final offer of about \$46.8 billion — \$95 a share — will buy it the 44% or so of Genentech that it doesn't already own. The merger is the latest of several mega-mergers rippling through the drug industry (see 'Getting it together').

But the Roche–Genentech deal has a unique resonance in the science and biotech communities, because Genentech's creative culture and rigorous science have been seen as essential to its ongoing success in producing novel medicines. And although Roche has vowed to preserve Genentech's entrepreneurial scientific culture, observers say that this may not be possible under Swiss leadership.

"It just seems unlikely that Genentech can maintain its creativity, because the difference between the two corporate cultures is not a superficial one," says Laurence Lasky, who worked at Genentech from 1982 to 2002 and was a Genentech fellow — the company's top science rank — when he left. "It's sad, because Genentech really is one of the greatest home-grown things that we have created in this country," says Lasky, who is now a partner at US Venture Partners of Menlo Park, California.

In a statement on the deal, the two companies said that Genentech's "unique culture [will be] maintained". Company executives are not speaking to the press about how exactly this will be accomplished. But former Genentech employees say that similar fears surfaced in 1990, when Roche bought a majority stake in Genentech, and proved to be unfounded. "That partnership didn't hurt our creativity; Roche stayed at a distance and Genentech continued to do great work," says David Goeddel, who was the first scientist hired at Genentech and is now a managing partner at the venture-capital firm the Column Group in San Francisco.

But, Goeddel adds, "I think the complete loss of any independence for Genentech is kind of sad after so many years".

Founded in 1976, Genentech grew from a tiny start-up to an almost \$100-billion company by capitalizing on pioneering science. For instance, Genentech scientists were the first to engineer bacteria to produce human proteins, such as insulin, which became the company's first drug — and the world's first genetically engineered drug — in 1982.

Since then, Genentech has developed a



P. SAKUMA/AP

Genentech's reputation for scientific creativity is often credited with its success in drug development.

pipeline of biological drugs worth \$9.5 billion in annual sales in the United States alone. Three of these are now Roche's best sellers. One, the cancer medicine bevacizumab, sold as Avastin, is already worth \$4.8 billion a year and is currently in clinical tests, due to report results in about a month, that may expand its uses.

Genentech's story has served as a model for the rest of the biotech industry, and young companies often speak of aspiring to become the next Genentech. But, Goeddel says, "that's pretty impossible to do now". Panicked by their own lack of promising drug candidates, big pharmaceutical companies are picking off small biotech at a much earlier stage than they did in the

past, preventing the companies from evolving independently. Some of the other early biotech companies have survived as a result of mergers and acquisitions: Biogen, based in Cambridge, Massachusetts, merged with Idec of San Diego, California, in 2003, and Amgen, headquartered in Thousand Oaks, California, has acquired several smaller biotech over the years.

Roche is thought to have been spurred to solidify its hold over Genentech because its exclusive rights to sell Genentech products outside the United States were set to expire in 2015. Roche first bid for Genentech last July, offering \$44 billion, or \$89 per share, for the rest of the company.

Now that the two firms have agreed on a final number, the merged company is set to become the seventh largest pharmaceutical firm in the United States, with an annual revenue of about \$17 billion. It will retain the Genentech brand in the United States and will be headquartered at Genentech's South San Francisco campus.

But the campus may see an exodus of Genentech employees, possibly including chairman and chief executive Arthur Levinson, who joined the company in 1980. "My guess is it would be very hard for [Levinson] to stay, because he's used to running his own show," Lasky says. Levinson is not speaking to reporters, but said in a statement: "We look forward to working with our partners at Roche to ensure a smooth transition once the transaction is complete and to continue our mission of serving patients." ■

Erika Check Hayden

## Getting it together

2009 is shaping up to be the year of the pharmaceutical mega-merger. Analysts say that such mergers are likely to be only a short-term fix for companies' dwindling drug pipelines. Among the recent highlights:

**January 2009:** Pfizer acquires Wyeth for \$68 billion, in part to make up for the fact that its blockbuster cholesterol drug Lipitor comes off patent in 2011.

**March 2009:** Merck buys Schering-Plough for \$41 billion, in part to acquire the company's pipeline of drugs in the late stages of development.

**March 2009:** Roche, which already owned a majority stake in Genentech, buys out the rest of the company for \$47 billion.

# Incoming chief to tackle woes of US food and drug agency

C. LIVINGSTON/GETTY IMAGES

Physician Margaret Hamburg, once the youngest-ever health commissioner for New York City, was nominated on 14 March by President Barack Obama to head the beleaguered US Food and Drug Administration (FDA). If confirmed by the Senate, she would bring to the commissioner's job a background of public-health experience that runs from working at the National Institute of Allergy and Infectious Diseases to advising the Centers for Disease Control and Prevention, and from bioterrorism planning to pandemic-flu preparation and tuberculosis control.

The president also named Joshua Sharfstein, a paediatrician who is the health commissioner in Baltimore, Maryland, to be Hamburg's principal deputy commissioner. His appointment does not require Senate confirmation.

The \$2.6-billion agency, still regarded as a global gold standard for drug regulation, has seen its image tarnished and its morale sapped in the past decade by a slew of regulatory failures related to both food and drugs (see *Nature* 457, 770–771; 2009). Numerous reports by government and scientific bodies have documented the erosion of the FDA's scientific and oversight capabilities, even as it faces burgeoning responsibilities. And episodes such as its delay in making the emergency contraceptive Plan B available over the counter led to charges of undue political influence on its decision-making under former commissioner Andrew von Eschenbach.

Obama, speaking in his weekly video address, said the FDA has been “underfunded and understaffed”, and announced the formation of a Food Safety Working Group of top government officials.

People familiar with Hamburg and Sharfstein said last week that although the two are not bench scientists, they would bring a science-based approach to the FDA.

Hamburg “has a great appreciation that public policy must be driven by the very best science”, says Gail Cassell, a microbiologist and vice-president of scientific affairs at drug company Eli Lilly, based in Indianapolis, Indiana. For instance, as health commissioner in New York in 1991, Hamburg campaigned vigorously for — and achieved — the establishment of a then-controversial needle-exchange programme to slow the transmission of HIV/AIDS. She also vocally opposed a requirement that AIDS educators stress sexual abstinence to city students.

In a rare move, Republican mayor Rudolph



**Margaret ‘Peggy’ Hamburg will lead the FDA.**

Giuliani asked Hamburg to stay on after he took over from Democrat David Dinkins, who had hired her. During her tenure, she boosted childhood immunization rates to record levels while dramatically lowering tuberculosis incidence by implementing a novel programme later cited as a model by the World Health Organization. She also created the country's first modern bioterrorism preparedness programme.

Hamburg left New York to become a lead strategist for the Department of Health and Human Services under then-president Bill Clinton, where she was in charge of planning for bioterrorism, among other things. “She was definitely ahead of the curve” in seeing the need for bioterrorism preparedness, says Thomas Inglesby, deputy director of the Center for Biosecurity of the University of Pittsburgh Medical Center in Pennsylvania. Most recently, she has served as a senior scientist at the Nuclear Threat Initiative in Washington DC, where she developed a strategic plan and grant-making portfolio addressing biological threats.

Sharfstein, a former staffer for liberal Congressman Henry Waxman (Democrat, California), has a history of confronting the drug industry. In 1997, as a medical resident, he wrote to *The New England Journal of Medicine* complaining that a Pfizer-sponsored event promoting its products to young doctors in Boston, Massachusetts, offered free pool and alcohol, urging them to “rack ‘em up & toss ‘em down”. (In a response, the company admitted that its slogan “created a poor impression”.)

In 2007, as health commissioner in Baltimore, Sharfstein successfully petitioned the FDA to oppose the use of over-the-counter cough and cold medicines in children under two years old.

**Meredith Wadman**



## Yale climate-change institute launched

Rajendra Pachauri, chairman of the Intergovernmental Panel on Climate Change (IPCC), will head a new climate and energy research institute at Yale University.

The institute and its head were announced on 10 March at an international climate-change conference in Copenhagen (see 'Soundbites from Copenhagen').

Pachauri, *Nature's* Newsmaker of the Year in 2007, will work part-time at the Yale institute, retaining his current positions with the IPCC and as director-general of India's Energy and Resources Institute in New Delhi.

## NIH research to become permanently open access

The annually renewable public-access policy first implemented by the US National Institutes of Health (NIH) a year ago was

made permanent in a spending bill signed into law on 11 March. The provision specifies that all NIH-funded investigators must submit an electronic version of their peer-reviewed manuscripts to the National Library of Medicine's PubMed Central. The manuscripts must be made publicly available no later than 12 months after publication.

The new law specifies that it must be implemented in a manner consistent with copyright law. Congressman John Conyers (Democrat, Michigan) is hoping to use that clause to upend the policy. His Fair Copyright

in Research Works Act would amend US copyright law to forbid federal agencies from making funding agreements, such as NIH grants, conditional on authors making copies of their peer-reviewed articles freely available.

## Maths mistake blamed for killing off 'clean coal' plant

The huge cost overruns cited as a reason for cancelling the US government's flagship carbon-sequestration plant, FutureGen,

## SOUNDBITES FROM COPENHAGEN

Views from the International Scientific Congress on Climate Change, 10–12 March.

**"On any kind of pragmatic timescale, I think we should see loss of the Amazon forest as irreversible."**

Chris Jones Met Office Hadley Centre, Exeter, UK

**"I'm frustrated, as are many of my colleagues, that 30 years after the US National Academy of Sciences issued a strong warning on CO<sub>2</sub> warming, the full urgency of this problem hasn't dawned on politicians and the general public."**

Stefan Rahmstorf Potsdam Institute for Climate Impact Research, Germany

**"Ice loss in Greenland has accelerated over the last decade. The upper range of sea level rise by 2100 might be 1 metre or more."**

Konrad Steffen University of Colorado, Boulder

**"Do the politicians understand just how difficult it could be, just how devastating 4, 5, 6 degrees centigrade would be? I think, not yet."**

Nicholas Stern former UK government economics adviser

For coverage of the conference, see <http://tinyurl.com/dm9plg> and <http://tinyurl.com/daroma>



E. SCHULZ/AP Rajendra Pachauri.

SOURCES: THE GUARDIAN, NATURE, UNIV. COPENHAGEN, THE TIMES

FERMILAB

were due in part to a mathematical error, a congressional report has found.

Announced by President George W. Bush in 2003, the plant was to have captured and stored 90% of its carbon dioxide emissions. It was originally slated to cost \$950 million, but was cancelled in early 2008 when new estimates pegged its cost at \$1.8 billion (see *Nature* 451, 612–613; 2008).

But those estimates were not comparable, the Government Accountability Office report says. The \$950-million estimate was done in constant dollars; the \$1.8-billion figure included expected increases for inflation. The 2008 figure would have been \$1.3 billion if adjusted for constant dollars.

Steven Chu, the new energy secretary, said last week that he plans to meet soon with private partners involved in the project to discuss whether and how to take it forward.

## Lighter Higgs boson harder to find

Physicists have placed fresh limits on the mass of the Higgs boson — the particle thought to confer mass on other matter.

The particle was assumed to have an energy (or equivalent mass) of between 114 and 185 gigaelectronvolts (GeV). But, after analysing pooled data, researchers



Fermilab's finding may cause problems for CERN.

at the Tevatron particle accelerator at the Fermi National Accelerator Laboratory in Batavia, Illinois, announced on 13 March that the particle can be excluded at energies of between 160 and 170 GeV, and is likely to exist at the lower end of the assumed range.

At more-crowded lower energies, filtering out other debris and finding a rare Higgs event becomes more difficult. Scientists at the Large Hadron Collider at CERN, Europe's particle-physics laboratory near

Geneva, had hoped to find the Higgs boson in the higher-energy region.

For a longer version of this story, see <http://tinyurl.com/cp8avk>.

## Hospital investigation reveals long-running fraud

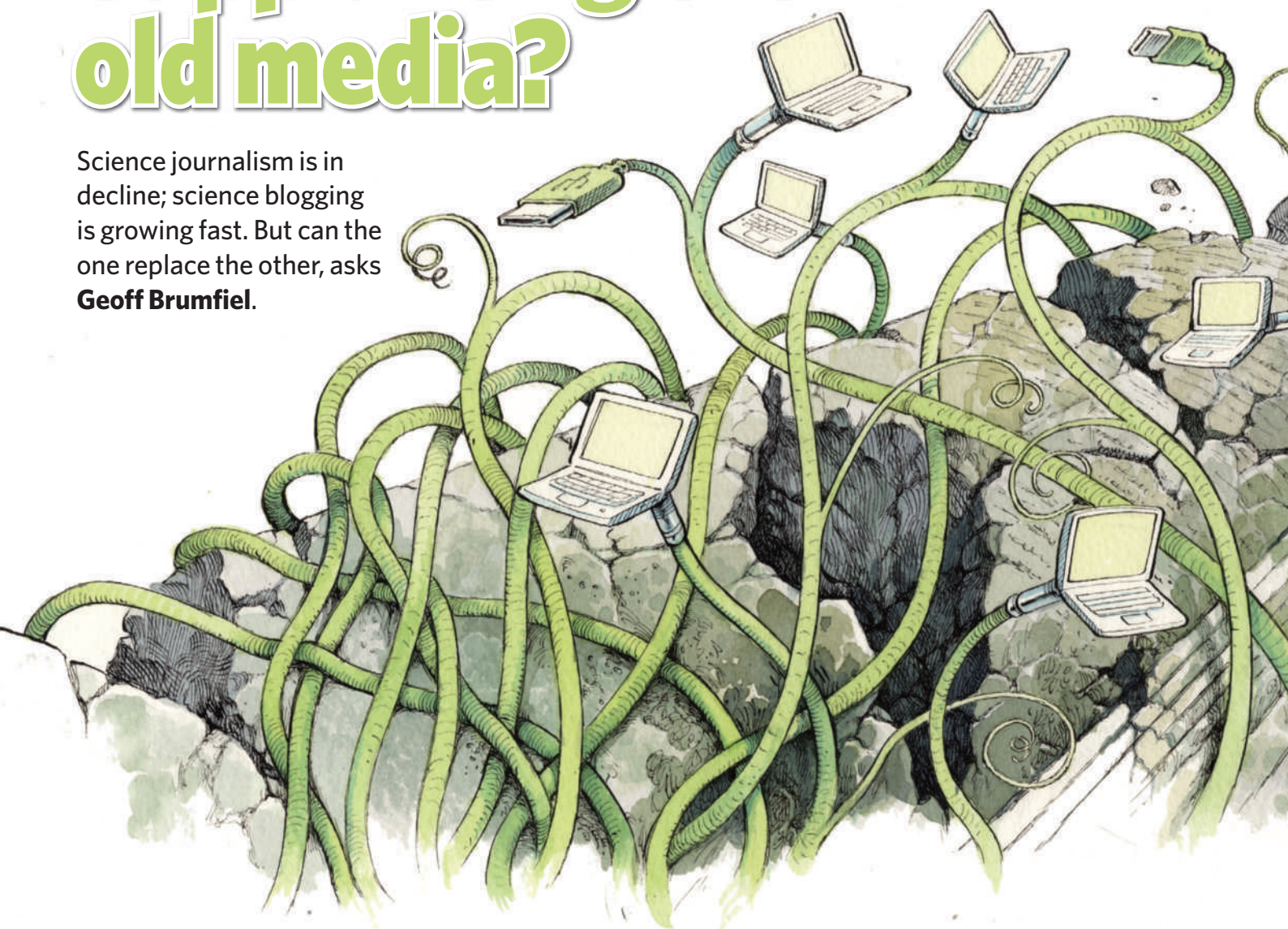
In what could be one of the largest ever cases of medical-research fraud, a prominent anaesthesiologist has been accused of fabricating data in at least 21 papers over 13 years.

Scott Reuben fabricated all or some of the data in studies as far back as 1996, according to an internal investigation by Baystate Medical Center in Springfield, Massachusetts, where he worked. Journals such as *Anesthesiology* and *Anesthesia and Analgesia* have already retracted Reuben's papers, which supported combining painkillers called COX-2 inhibitors (such as Merck's Vioxx and Pfizer's Celebrex) with other analgesics, including Pfizer's Lyrica, to relieve pain after surgery.

An attorney representing Reuben said that the Baystate investigation was confidential, that there were extenuating circumstances and that Reuben deeply regrets all that has happened. Reuben is barred from research at Baystate for at least a decade.

# Supplanting the old media?

Science journalism is in decline; science blogging is growing fast. But can the one replace the other, asks **Geoff Brumfiel**.



**J**ohn Timmer's slide into journalism was so gradual even he can't put his finger on the point at which he stopped being a researcher.

He started reading Internet websites and message boards a decade ago, while he was working as a postdoc in a developmental neurobiology lab at the Memorial Sloan-Kettering Cancer Center in New York. One day, one of his favourite sites, *Ars Technica*, announced that it was looking for someone to help with its science coverage. It was 2005, and a school board in Dover, Pennsylvania, had gone to court over the promotion of intelligent design. "I thought, wow, it really feels like the public has completely lost touch with what science is all about," says Timmer. "So I basically e-mailed the existing author and volunteered."

Over the next few years Timmer's work on the site grew steadily, while his research

career stalled. Today the 42-year-old draws a full-time salary as *Ars Technica*'s science editor. He works with writers echoing his earlier experience: graduate students and postdocs type up brief summaries on research in their areas of expertise during down time and lunch breaks. The write-ups are more technical than you might read in a newspaper — a recent post included a lengthy discussion on 'functionalizing' cells to bind them together with DNA — but that's fine, Timmer says. The idea is to provide people already interested in science with greater insight into how research works. A typical posting can earn a writer anywhere from the price of a pair of movie tickets to around US\$100, and that is often incentive enough for young academics.

Timmer's tale is emblematic of a shift in the way science meets the media. In part because of a generalized downturn, especially in

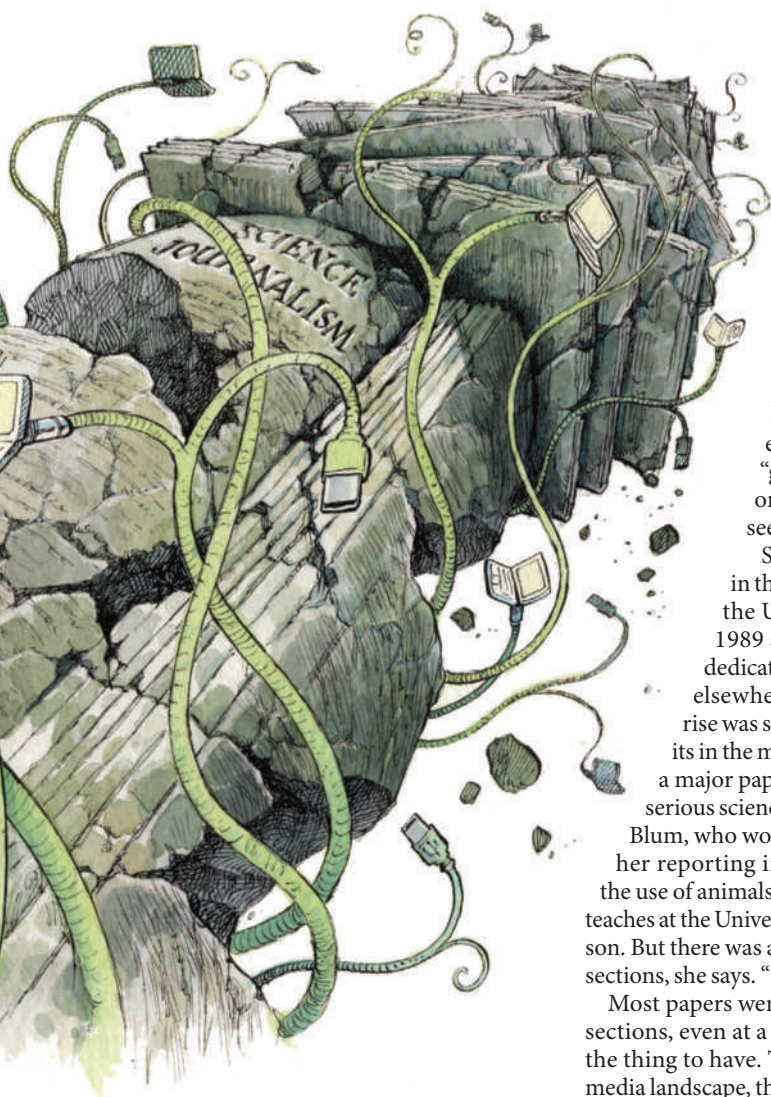
newspaper revenues, the traditional media are shedding full-time science journalists along with various other specialist and indeed generalist reporters. A *Nature* survey of 493 science journalists shows that jobs are being lost and the workloads of those who remain are on the rise (for full results see <http://tinyurl.com/c38kp6>). At the same time, researcher-run blogs and websites are growing apace in both number and readership. Some are labours of love; others are subsidized philanthropically, or trying to run as businesses.

## It's a blog world

Traditional journalists are increasingly looking to such sites to find story ideas (see 'Rise of the blogs', page 276). At the same time, they rely heavily on the public-relations departments of scientific organizations. As newspapers employ fewer people with science-writing

DAVID PARKINS





backgrounds, these press offices are employing more. Whether directly or indirectly, scientists and the institutions at which they work are having more influence than ever over what the public reads about their work.

The amount of material being made available to the public by scientists and their institutions means that “from the pure standpoint of communicating science to the general public, we’re in a kind of golden age”, says Robert Lee Hotz, a science journalist for *The Wall Street Journal*. But that pure standpoint is not, or should not be, all that there is to media coverage of science. Hotz doubts that blogs can fulfil the additional roles of watchdog and critic that the traditional media at their best aim to fulfil. That sort of work seems to be on its way out. “Independent science coverage is not just endangered, it’s dying,” he says (see ‘Vox media’, page 277).

What’s more, the amount of material available is not a good proxy for its reach. Press releases and blogs will not find the same broad

audience once served by the mass media, says Peter Dykstra, who was executive producer of CNN’s science, technology, environment and weather unit until it was closed down last year. Now at the Woodrow Wilson International Center for Scholars, an independent think tank in Washington DC, he says that science and environment news will be “ghettoized and available only to those who choose to seek it out”.

Science journalism boomed in the 1980s and early 1990s. In the United States — where by 1989 some 95 newspapers had dedicated science sections — and elsewhere, the field’s precipitous rise was supported by buoyant profits in the media sector. “The model of a major paper was that they did really serious science coverage,” says Deborah Blum, who won a 1992 Pulitzer Prize for her reporting in the *Sacramento Bee* on the use of animals in research, and who now teaches at the University of Wisconsin at Madison. But there was a problem with the science sections, she says. “They didn’t make money.”

Most papers were willing to support their sections, even at a loss, because science was the thing to have. Today, in a harsher mass-media landscape, that has changed. Across the United States, newspaper science sections have been shut down: this month *The Boston Globe* stopped running its weekly science and health section. Nor is the written word the only casualty, as the closure of Dykstra’s seven-person unit at CNN indicates. *Nature*’s survey shows that, of those working in the United States and Canada, one in three had seen staffing cuts at their organization (see ‘Hiring practices’).

The European industry has not yet reached the level of crisis seen in the United States, says Holger Wormer, a professor of science journalism at the University of Dortmund in Germany. Many newspapers in Germany are considering staff cuts but, at the moment, science journalists are faring relatively well. “Science departments are still small but they are regarded as quite important,” he says. Because larger German papers such as *Frankfurter Allgemeine Zeitung* have science sections, smaller papers are willing to support their own science coverage, at least for now. In France, declining circulations are also creating problems, according to Stéphane

Foucart, a science writer at *Le Monde*. In the past six months, *Le Monde* has scaled back its science coverage. Newspapers and broadcast outlets in the United Kingdom are also under pressure, and science and environmental jobs are among those that have been lost.

Unsurprisingly, among the science reporters who remain, the workload is on the rise. *Nature*’s survey reveals that 59% of journalists have seen the number of items they work on in a given week increase over the past five years. They are not just doing more reporting, but more types of reporting. Many are now being asked to provide content for blogs, web stories and podcasts — something they weren’t doing five years ago.

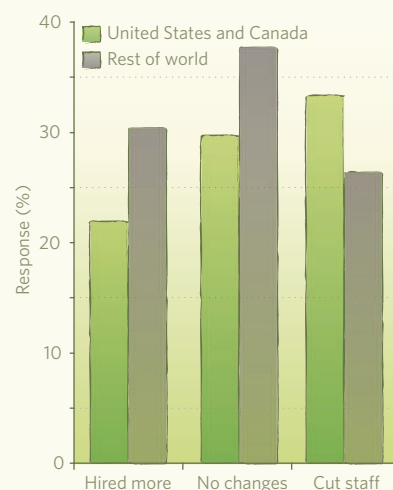
### Fast and dirty

Under these straitened conditions the mainstream media’s need for quick and accurate science content is being met primarily by public-relations departments, according to Fiona Fox, director of the Science Media Centre, an organization in London that supplies journalists with scientific information (*Nature*’s editor-in-chief, Philip Campbell, sits on the Science Media Centre’s board, and the Nature Publishing Group provides support for it). Mark Henderson, science editor for *The Times*, based in London, says that he tries to avoid relying solely on releases “as much as possible”, but “if there’s a good press release and you’ve got four stories to write in a day, you’re going to take that short cut”. *Nature*’s survey shows press releases to be a top source of story ideas for science journalists, with 39% routinely quoting from them directly.

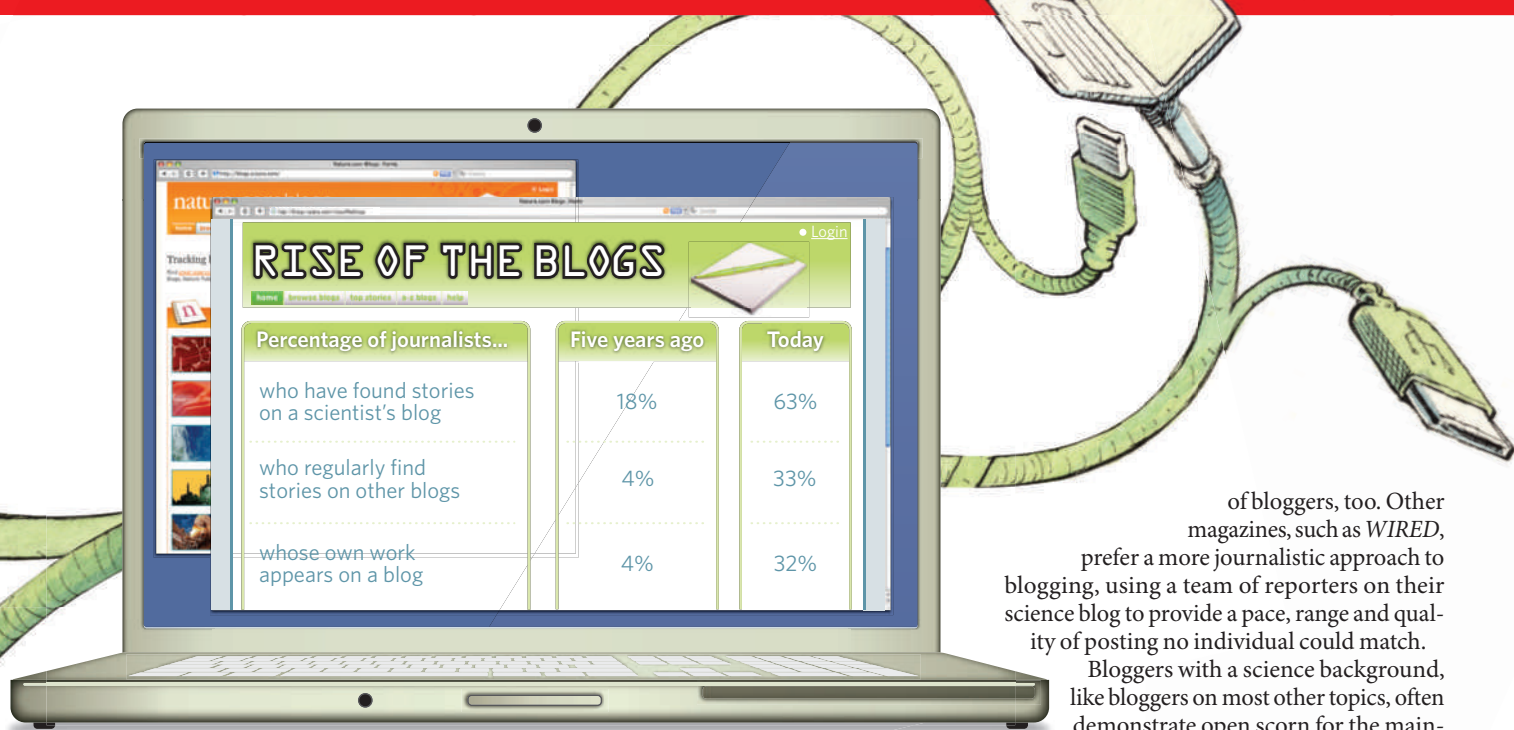
This demand for stories and ideas has been matched by an increase in supply. In Britain as

### HIRING PRACTICES

Many North American science journalists report job losses in the past five years.







in the United States, contraction in the media has made jobs in public relations particularly attractive for students at science-writing programmes. "You'd be amazed at the diversity of places for science communicators," says Blum. Government agencies, universities, museums and non-governmental organizations have all hired her students, she says — almost all of whom are finding jobs, despite the woes of the traditional media.

The Science Media Centre demonstrates the new opportunities that exist now. It was started in 2002 by an amalgam of non-commercial and commercial interests seeking to influence the public debate on news topics such as genetically modified foods. What began as a relatively modest attempt to connect journalists to sources of scientific expertise has expanded dramatically over the past seven years. Today, the centre's six-person staff sends out daily e-mails filled with quotes from prominent scientists on the latest news that end up in tomorrow's stories. It has also begun providing fact boxes and background documents that journalists can insert directly into their coverage. Fox is happy at the centre's success, but uneasy too. Ideally, she says, science journalists should be picking up the phone and talking to scientists directly: "We are successful because of a serious problem in journalism, and it's not one to be celebrated."

### Straight to the masses

As journalists become more dependent on scientific public relations, scientists themselves have begun reaching out to mass audiences through the Internet. Such outreach is not new; but unlike books and lectures, science blogs operate with a quick turnaround that more closely resembles that of the traditional media. The most successful sites are drawing hundreds of thousands of visitors each month.

Many of those blogs were started by scientists who simply wanted to reach the public with information about their research. "I'd always find that people were interested in what I did," says Derek Lowe, a researcher with Vertex Pharmaceuticals in Cambridge, Massachusetts, and author of *In the Pipeline*, a blog about drug discovery and the pharmaceutical industry. "Most people have no idea how drugs are actually found," he says. Lowe started his blog in early 2002, and now it regularly draws around 200,000 page views a week.

Paul Myers, a biologist at the University of Minnesota in Morris, says that he started his blog *Pharyngula* "largely out of boredom", but now that he gets more than half-a-million weekly page views, he sees it as a valuable tool for talking to a public audience. Myers freely admits that his readers "are not just there for the science" — his attacks on religion are a mainstay of the blog's appeal. But he certainly considers himself a source of scientifically reliable information for his readers.

Although science blogging did not start off as a business, there are attempts to make it one. Since 2006, the publisher of *Seed*, a magazine about science, has gathered more than 100 science blogs — including *Pharyngula* — on a range of topics on to a single website, *ScienceBlogs*, and now pays its bloggers on the basis of how many hits their posts receive. Fabien Savenay, a senior vice-president for marketing at *Seed Media Group* in New York, declines to say whether the blog site makes money for the organization. But, he says, the project "has been a successful franchise for us in that it has great traffic and engagement". Another US magazine, *Discover*, has recently been amassing a smaller but impressive stable

of bloggers, too. Other magazines, such as *WIRED*, prefer a more journalistic approach to blogging, using a team of reporters on their science blog to provide a pace, range and quality of posting no individual could match.

Bloggers with a science background, like bloggers on most other topics, often demonstrate open scorn for the mainstream media (MSM in blogpeak).

"You get a press release that is slightly rehearsed by somebody in the newsroom and it goes in the paper! It's wrong, its sensationalist, it erodes the public trust in scientific endeavour," says Bora Zivkovic, author of *A Blog Around the Clock* on *ScienceBlogs* and an online community manager for the Public Library of Science journals. Myers takes a similar view. "Newspapers realize that they can get their audience by peddling crap instead of real science," he says. Not surprisingly, those who came to blogging from journalism — such as Carl Zimmer, who writes for a range of publications, including *The New York Times*, and blogs at *Discover* — tend to disagree. But Larry Moran, a biochemistry professor at the University of Toronto, Ontario, who blogs at *Sandwalk*, seemed to speak for many bloggers

when he recently wrote "Most of what passes for science journalism is so bad we will be better off [sic] without it".

While journalists such as Zimmer expand their

mainstream work into their blogs, bloggers with roots in the lab are moving into print. Myers will soon contribute a regular column to the *Guardian* newspaper in the United Kingdom. Derek Lowe now writes regular columns for *The Atlantic* and the trade magazine *Chemistry World* (both have also written for *Nature*). This work, though, tends towards opinion and analysis, not reporting. "Bloggers don't want to be journalists," says Zivkovic. "I want to write on my blog whatever I want. I may write a post about a new circadian paper, but the next eighty posts are about politics or what I ate for breakfast." Despite his distaste for how the trade is practised, he thinks that there will always be a need for professional journalists covering science. "Somebody has

"It feels like the public has completely lost touch with what science is all about."

— John Timmer

## Vox media

More than 100 science journalists responding to *Nature's* survey offered their thoughts on the future of the field. Here's a sample of what they had to say:

**"Science journalism is dying in the mass media. It has always been a niche subject, but only those really interested in it will continue to purchase specialist science media. Print publications will become more niche but will survive. TV news and documentaries will become dumbed down in order to compete with the idiocy on the Internet."**

"The public remains interested in science. They pack science fairs and museums; they buy popular science books; they watch TV documentaries. But I'm not sure the public's appetite for science is so great that people need daily science news. So when this or that media outlet cuts its science desk, it could be in response to what they can now measure on their websites: which topics really engage the public day to day."

**"I am a scientist who is freelancing occasionally for a science popularization magazine published by my institute. Most of the time, the description of the scientific result in a press release is so dumbed down that I cannot find out what the result actually was in the terms of an expert! Instead of dumbing down the science to the level of the general public, we should be trying to educate the public."**

"It has been shocking to see the public come to view science news as a bulk commodity. Readers seem to make little or no distinction between professionally written reports from independent news organizations and promotional writing masquerading as news on various blogs and science 'news' websites."

**"Commercial pressures are polluting science journalism. The mainstream media has pitifully low standards of science journalism where the herd mentality prevails. There is a prevailing view among newspaper editors that science does not deserve as much coverage as other fields, founded probably on nothing other than these editors' personal chip on their shoulder regarding their own scientific education."**

"I'd love to know if the monks were wringing their hands over the horrible shallowness of thought sure to follow the invention of those funny little letter bits squashed on paper with a press."

G.B.

to actually be paid to write about things as they come out," he says.

That is what John Timmer is looking for new ways to do at *Ars Technica*. But there is a problem: the online world, both in its bloggier reaches and elsewhere, is polarized; people go to places they feel comfortable. Many of the people that Timmer originally hoped to reach when writing about intelligent design and the Dover trial probably go elsewhere for their news, he says, because "it's easy for somebody to pick their news sources based on their politics, and get that version of scientific issues". Dykstra worries that in a more fragmented media world, "environmental news will be available to environmentalists and science news will be available to scientists. Few beyond that will pay attention."

Others worry about the less questioning approach that comes with a stress on communication rather than journalism. "Science is like any other enterprise," says Blum. "It's human, it's flawed, it's filled with politics and ego. You need journalists, theoretically, to check those kinds of things," she says. In the United States, at least, the newspaper, the traditional home of investigations and critical reporting, is on its way out, says Hotz. "What we need is to invent new sources of independently certified fact."

### Culture mash

Two Ivy League giants, Princeton University in New Jersey and Yale University, are trying to do something about the problems they see in environmental coverage with websites aimed at generating scientifically accurate news coverage. "We're bringing something new to the table," says Roger Cohn, a veteran journalist who now edits the Yale Environment 360 website, which is funded in part by the William and Flora Hewlett Foundation and the John D. and Catherine T. MacArthur Foundation. The site is home to reports by journalists and opinions by scientists on subjects such as climate change, but it has "no axe to grind on any one of these issues," says Cohn.

At the Princeton University website, Climate Central, the focus is mainly on video material. "We're just in the initial stages of preparing a weekly series of news stories about climate based on papers in journals," says Michael Lemonick, a long-time science

writer for *Time* magazine who now works at the site. As well as appearing on Climate Central, he says, the stories will be offered to the websites of big media outlets; some of the group's work has already been aired on the Public Broadcasting Service's evening news show *The NewsHour with Jim Lehrer*, which reaches millions of viewers. Climate Central is funded by

the Flora Family Foundation and The 11th Hour Project, a non-profit organization supporting climate awareness, based in Palo Alto, California.

Lemonick says his new job requires him to listen more closely to research-

ers. "If they say, 'you really left out this important fact,' I don't get to say, 'Sorry it's my story,'" he says. That doesn't mean that researchers make his story into a dry scientific paper, he adds. "They have to recognize the needs of the journalist, but we have to recognize the needs of the scientists. We're kind of fusing the two cultures." Timmer's path has also led him to a fusion of science and journalism. In May, media giant Condé Nast acquired *Ars Technica*, and he was brought on full-time. "When I'm interacting with press officers or researchers, I'm acting as a journalist," he says. "I don't think anybody would consider me a working scientist any more." But when asked how he sees the scientists writing for him, he becomes more philosophical: "Basically, however they see themselves."

Geoff Brumfiel is a senior reporter in *Nature's* London office.

See Editorial, page 260.

Full survey data accompany this article online.

**"Science is just like any other enterprise. It's human, it's flawed, it's filled with politics and ego."**

— Deborah Blum



DAVID PARKINS





# Archived answers

The switch to electronic medical records opens up a potential wealth of data for researchers, if major obstacles can be overcome, reports **Katharine Gammon**.

In May 1999, the US Food and Drug Administration (FDA) approved a promising new drug. Earlier clinical trials of 5,000 patients had convinced regulators that the drug was a safe and effective means of reducing pain and inflammation in arthritis. Produced by the pharmaceutical giant Merck, the new medication became an instant hit and nearly 80 million people worldwide had taken this wonder pill by 2004.

During that same period, however, evidence emerged that the painkiller, known as Vioxx (rofecoxib), significantly raised the risk of heart attacks. By the time Merck pulled the drug from the market in September 2004, “an estimated 88,000 to 140,000 excess cases of serious coronary heart disease probably occurred in the USA”, according to a study led by an FDA scientist and published in the *Lancet* in 2005 (ref. 1). The drug may have caused tens of thousands of deaths in the United States alone, the researchers estimated.

The digital revolution — and the switch from paper to electronic medical records — might prevent many of these kinds of deaths

in the future. In a 2007 study<sup>2</sup>, researchers from the Children’s Hospital Informatics Program (CHIP) in Boston, Massachusetts, attempted to see how quickly they could have spotted a problem with Vioxx using digital records. By homing in on two major hospitals, the researchers found an 18% jump in heart attack cases within eight months of Vioxx being introduced, which could have provided an early warning about the drug, says Isaac Kohane, an author of the study and director of CHIP. The excess cardiac cases disappeared when Vioxx was pulled from the market.

The Vioxx incident shows one benefit that could emerge when health systems dump old-fashioned paper medical charts in favour of electronic records. Some countries, such as the United Kingdom, have already made the transition and the United States is moving in that direction. This trend could have profound implications for how doctors treat patients and how researchers pursue

scientific questions that rely on medical data.

“In the long run, the opportunities are huge — actually transformative,” says Ashish Jha, assistant professor of health policy at the Harvard School of Public Health in Boston. “When the exchange of health information flows freely between doctors, we can’t imagine how huge that’s going to be.”

But it will take time. The move towards electronic medical records has proceeded slowly so far, and many practical challenges still lie ahead. When Jha and his colleagues surveyed US physicians in late 2007 and early 2008, they found that just 4% have a full electronic records system, and only 13% have even a basic digital system<sup>3</sup>. “The organization and delivery of health care hasn’t changed in nearly a hundred years,” says Jha.

That may soon change. The stimulus package that recently passed in US Congress contains US\$19 billion to promote the switch to digital health records, mostly through economic

**“The organization and delivery of health care hasn’t changed in nearly a hundred years.”**  
— Ashish Jha

R. ESTAKHRIAN/GETTY

incentives for doctors and hospitals to make the transition. The lofty goal is to have all health records stored and accessed electronically by 2014.

President Barack Obama has promoted this policy as a way to control soaring health-care costs. An estimated \$81 billion could be saved annually if 90% of doctors used electronic systems — by reducing redundant care, speeding up patient treatment and stemming mistakes and handwriting foul-ups. And the switch could cut more than just costs. By reducing the number of medication errors, electronic systems can save lives.



Digital health records can save lives and money, but security is a concern.

### Digging for data

Proponents of electronic medical records say that researchers could be among the chief beneficiaries of the conversion to digital health data. The fundamental argument for using electronic records in scientific studies is that most doctors don't see enough patients to recognize patterns. Only on the level of hundreds of thousands or millions of records do connections become apparent. What's more, research can be slow when it comes to recruiting patients, evaluating them and checking in over a long period of time, but all that information is available just a few clicks into an electronic record.

But critics argue that current health files are too unstructured, too unreliable and that algorithms for sifting doctors' notes are not yet available.

There's also the question of the type of information in the record. "Depending on what you want to study, that may or may not be consistently documented," says Samuel Skootsky, medical director of the Medical Group and Faculty Practice Group at the University of California, Los Angeles.

Kohane counters that there is enough valuable information in records to justify mining them. What's more, the gigantic amount of patient data available means researchers can choose the individuals to study. For example, investigators looking at asthma could pluck out true asthmatics from people who wheeze owing to other factors, he says.

For research purposes, he says, observations in medical records more accurately reflect clinical practice. "If you study a drug and you look at its effect on a control group, that's a very pure study," says Kohane. "But everyone taking that drug might be also taking two other drugs related to their condition", which could alter the effects of the medication being studied. That information is not always available

to researchers conducting trials, but electronic medical records might help sort out the multi-drug effects.

In a test of that potential, Kohane's group is using health records to look at other drugs on the market. The team is searching for effects that might have been missed in earlier tests of the medications.

Even more power might emerge by combining clinical records with a genetics database. Michael Swiernik, director of medical informatics at the University of California, Los Angeles, says that the combination would allow researchers to find the right people for a study in a speedy, cost-effective way.

Other research affirms that it is possible to dig into electronic records for useful clinical data. A study released in February showed that data captured in electronic records could yield the same results as data from clinical trials when evaluating the efficacy of a drug<sup>4</sup>. The data came from Britain's general practice research database, which holds about 10% of the records of general practitioners — mostly relating to demographics and pharmaceutical usage. The problem is that these UK databases are small, incomplete and are not yet linked together.

Before researchers can truly tap digital records, the databases first have to be created and filled. That will not happen if the primary users reject them, as happened at the Cedars-Sinai Medical Center in Los Angeles in 2002. After undergoing months of training with a new electronic records and communications system, doctors complained that it was too cumbersome. One day, the staff revolted, ending the hospital's \$34-million effort to convert to an electronic system after just three months.

half the doctors in the United States practise either alone or with one partner, the cost is extreme: \$32,000 per physician for installation and \$1,500 per month for running the system, according to estimates.

When a hospital or doctor goes out to buy a system, they face another challenge: there are hundreds of vendors who each want to sell a different, proprietary product, many of which will not survive. In addition, medical records

systems often have problems sharing data with each other. Taken together, these issues "make doctors and hospitals very nervous to purchase the technology", says Jha.

That apprehension can extend to using the systems as well: studies show that doctors'

efficiency actually drops 10–20% during the first year of using an electronic system.

Then there's the thorny issue of privacy. Moving electronic patient information around between doctors, pharmacies, laboratories and billing departments means many people have access to sensitive information. That raises the risk of a large security breach. Medical-record information sells for more on the black market than credit-card information, because it is more identifying, says Swiernik.

"Security is only as good as its weakest link, and the likelihood of weak links is really pretty high," says Lee Tien, a lawyer at the Electronic Frontier Foundation, a digital-rights public-interest group based in San Francisco, California. He is particularly concerned about data security because the transition to electronic health records is happening so quickly. "If you're not hard-wiring privacy and security into the DNA of the infrastructure, there will be unfortunate data security breaches."

Despite these challenges, some institutions

**"What you want to study may or may not be consistently documented."**

— Samuel Skootsky

L. LEFKOWITZ/GETTY



are successfully using and benefiting from electronic medical records — especially large organizations that act as the payer as well as the provider of health care. Two of the largest systems are VistA, the open-source system of the Department of Veterans Affairs (VA), and Kaiser Permanente's system HealthConnect. Kaiser Permanente's non-profit health plan has 8.6 million members with full electronic medical records, and nearly half of them use the online interface, where they can view part of their records.

But those systems have yet to benefit researchers substantially. Because of privacy regulations, Kaiser Permanente, based in Oakland, California, limits access to its clinical records mostly to researchers within the organization. The VA has also been slow in research. "The VA has had electronic data for 10 years now, but only a small amount of stuff is available for digital clinical research," says Jha, adding that the technology and the level of comfort with anonymized data sharing hasn't yet caught up to the amount of data available.

The largest electronic medical-records system in the world is Britain's National Health Service (NHS), serving more than 60 million people. But the digital data amassed by UK doctors is not yet comprehensively linked up. That will change in the next five years, by which time electronic records will connect 30,000 general practitioners and include data on all UK residents.

Despite that push into the digital age, the opportunity for research has been missed, says John Powell, associate professor in epidemiology and public health at the University of Warwick, UK. "There's a massive debate going on right now about consent — whether patients are opting in or opting out of a research system. Doctors for the most part are arguing that people have to give consent to opt in, civil servants in the Department of Health are saying they have to opt out. From the research point of view, we want to have maximum people and maximum data in." General practitioners tend to be wary of sharing patient data, even if they are de-identified.

That concern grew recently, when some officials suggested that the NHS could sell data to pharmaceutical companies to raise funds — a proposal that prompted strong

criticism. Powell says many of the issues will be resolved this year, when the NHS finalizes its official constitution — a sort of terms-of-service agreement for using the system. He argues that the system should be used for research. "In return for getting completely free health care for everything, there are certain things you have to sign up for as a resident — and one of those things should be secondary use of your data for the anonymous NHS research, but not for pharmaceutical companies," says Powell.

### Security fears

Britain has faced major data security problems in the past few years, so people are understandably concerned about the safety of their medical information. But there are rumblings of change: the creation in February of the National Institute for Health Research, a framework agency within the NHS, should make it easier for anonymized data to move within the system.

Other European nations haven't decided what to do with their records, either. In Germany, doctors have local electronic health records, but no way to connect them to a larger system. France has a national system to hold and access electronic medical records, although clinicians can view only the part of the record that is relevant to them.

Some researchers see an alternative to large, expensive systems in something closer to Apple's iPhone, with its panoply of available applications. Kohane says that the best system would be a platform available with different inter-operable, substitutable programs.

If open-source platform systems aren't built because of the nature of the vendors, Kohane sees a future in the expansion of personal health records — computerized records maintained

intermediaries — even offering the anonymity required by most review boards.

Such systems could fix a major problem in medical research. Right now, people who participate in studies do so anonymously, and the results are not usually transmitted back to them. Kohane and his colleagues are testing an approach later this year that will allow for more communication in one Boston clinic with 5,000 patients, and they are adding more clinics in 2010. Patients will sign up at their health-care institution and upload clinical or genomic information to an encrypted personal health record to be shared with researchers. Then patients could choose to 'tune in' to a certain type of announcement regarding their records.

If results came out that matched their genomic profile, patients would get a message about that work. They could also sign up for getting news on particular conditions, such as cancer or diabetes. And researchers could communicate with patients without knowing their identities, allowing an ongoing relationship. "The current state of research is of self-inflicted mutual ignorance," says Kohane. "Patients are committing to not knowing more about the study, and researchers can't find out more information about the patients."

All that could change if health systems can get past the substantial hurdles. In a world where information is at our fingertips, and even the US president has his thumbs on a BlackBerry, electronic medical records are coming, along with the research opportunities they hold. For many scientists, that transition is just what the doctor ordered.

**Katharine Gammon is a freelance writer based in Santa Monica, California.**

1. Graham, D. J. et al. *Lancet* **365**, 475–481 (2005).
2. Brownstein, J. S., Sordo, M., Kohane, I. S. & Mandl, K. D. *PLoS ONE* **2**, e840 (2007).
3. DesRoches, C. M. et al. *N. Engl. J. Med.* **359**, 50–60 (2008).
4. Tannen, R. L., Weiner, M. G. & Xie, D. *Br. Med. J.* **338**, b81 (2009).

**See Editorial, page 259.**



Consumers are storing their medical data online with systems such as Google Health.

**"If you're not hard-wiring privacy into the DNA of the infrastructure, there will be data security breaches."**

— Lee Tien

## CORRESPONDENCE

These letters respond to the News Feature 'Closing arguments' (*Nature* **457**, 650–655; 2009), about problems faced by laboratories funded by the US National Institutes of Health.

## Diversity of funding sources and topics is key to survival

SIR — Reduced funding and increased competition among a larger body of active researchers also affects countries other than the United States, and fields other than the biomedical sciences.

Here in the United Kingdom, those of us who use the Engineering and Physical Sciences Research Council (EPSRC) for funding have witnessed a similar scenario to that described in your News Feature: a smaller fraction of assessed grants are being funded and we have reached the stage where there is — for all practical purposes — a lottery to determine which of the many excellent proposals submitted will be successful.

The amount of money that is available for curiosity-driven research is also decreasing, as the government seeks to steer funding into areas of political priority. The response from the scientific community is, of course, to become more inventive about the potential applications of a given proposal.

I believe that the best way to maintain research activity is through diversity in both sources of funding and topic. You indicate that undue weight is given to National Institutes of Health funding (as opposed to general funding) in many US institutions as a requirement for tenure or promotion. This may further concentrate academics in attempting to access a finite pot, with a consequently reduced rate of success.

However, despite the disappointment that we may feel in not receiving enough funding to maintain our activity or research-group size at a desired level, we academics should recognize our privileged tenured

position in times of economic hardship such as the present.

**Brian Derby** School of Materials,  
University of Manchester, Grosvenor  
Street, Manchester M1 7HS  
e-mail: [brian.derby@manchester.ac.uk](mailto:brian.derby@manchester.ac.uk)

## Grant-writing offices would let scientists get on with research

SIR — The present financial crisis is putting many early-career scientists in a complicated situation, as noted in your News Feature and the Editorial 'Crisis of confidence' (*Nature* **457**, 635; 2009). I agree with the Editorial that "doctoral programmes should build better career counselling and training into their curricula from the start", but I also believe that universities cannot continue to expect professors to do research and apply for grants and patents, as well as teaching.

People often choose an academic career for the freedom and creativity of doing research. Overburdening them with administrative tasks means that their creativity is hardly used. One solution might be a 'grant-writing' office that would know all about the availability of grants, how they might apply to specific topics and what they should contain.

**Cecile Perrault** Department of  
Biomedical Engineering,  
McGill University, Montréal,  
Québec H3A 1A4, Canada  
e-mail: [cecile.perrault@mcgill.ca](mailto:cecile.perrault@mcgill.ca)

## We need more insight into what's worth paying for

SIR — An economic argument can be made that growth depends on the creation of new markets being opened up by new technology. Personalized medicine provides an example of this, as do advances

in medical instrumentation such as magnetic resonance imaging and laser eye surgery.

Balancing the need for increased spending to further economic growth with the need to spend more effectively during this economic downturn means that we must improve our ability to decide what research is worth funding. Simply lowering the percentage of funded grants won't cause this to happen by itself.

Effective spending depends on detailed peer review to assess applications properly. But the review quality necessary may be hard to realize in practice, because there is more and more research to review. The traditional rankings devised from impact factors don't take this kind of work into consideration — rather, they promote more writing, creating more work for reviewers. A better system would properly credit people for their review and assessment work so that money could be spent more effectively.

**William Gunn** Genalyte Inc,  
11760 Sorrento Valley Road,  
San Diego, California 92121, USA  
e-mail: [william.gunn@gmail.com](mailto:william.gunn@gmail.com)

## Backlogged system in Australia shuts out new investigators

SIR — The US situation may be worse than it once was. But it's still a big step ahead of Australia. Total health-related R&D spending as a percentage of gross domestic product is roughly half that of the United States or the United Kingdom, according to the Organisation of Economic Co-operation and Development (*Statistics and Indicators for 30 Countries* OECD, 2007). Our biomedical fellowship scheme is backlogged to the point that the awards designed to support new principal investigators go mostly to well-established group leaders.

A recent survey of the medical research workforce in Australia highlights the level of anxiety and discontent over poor career-development opportunities and lack of funding (M. Kavallaris *et al. Med. J. Australia* **188**, 520–524; 2008). In the preceding five years, for example, 6% of respondents had already left active research; 73% were considering leaving. This presents a clear challenge for recruitment and retention of a highly skilled workforce.

To keep Australian science at the leading edge, we need to maintain the flow of fresh ideas by ensuring opportunities for both new and established investigators. The new US administration has recognized the loss of career-development opportunities caused by underfunding. Australia runs the risk of failing to capitalize on significant government investments made so far. The recent announcement of a federal fellowship scheme for mid-career researchers hints at a potential improvement. But the near-term prospects for aspiring new investigators and those returning from overseas are not promising.

**Darren Saunders** Department of  
Pathology & Laboratory Medicine,  
University of British Columbia; and  
Molecular Oncology & Breast Cancer  
Program, BC Cancer Research Centre,  
Vancouver, BC V5Z1L3, Canada  
e-mail: [dsaunders@bccrc.ca](mailto:dsaunders@bccrc.ca)

*In February 2009, the director of the US National Institute of Neurological Disorders and Stroke moved the institute's funding cut-off for grant applications from the 10th to the 11th percentile after a re-evaluation of financial obligations. The change meant that Darcy Kelley (quoted in this News Feature) and another seven investigators whose applications had fallen below the cut-off in recent funding rounds were funded — Editor, Nature.*

Further discussion is welcome at  
<http://tinyurl.com/cf59fq>.



# ESSAY

## Do nations go to war over water?

**Wendy Barnaby** was asked to write a book about water wars — then the facts got in the way of her story.

The United Nations warned as recently as last week that climate change harbours the potential for serious conflicts over water. In its World Water Development Report<sup>1</sup> of March 2009, it quotes UN Secretary-General Ban Ki-moon noting the risk of water scarcity “transforming peaceful competition into violence”. It is statements such as this that gave birth to popular notions of ‘water wars’. It is time we dispelled this myth. Countries do not go to war over water, they solve their water shortages through trade and international agreements.

Cooperation, in fact, is the dominant response to shared water resources. There are 263 cross-boundary waterways in the world. Between 1948 and 1999, cooperation over water, including the signing of treaties, far outweighed conflict over water and violent conflict in particular. Of 1,831 instances of interactions over international freshwater resources tallied over that time period (including everything from unofficial verbal exchanges to economic agreements or military action), 67% were cooperative, only 28% were conflictive, and the remaining 5% were neutral or insignificant. In those five decades, there were no formal declarations of war over water<sup>2</sup>.

I learned this the hard way. A few years ago, I had just written a book about biological warfare<sup>3</sup> and the publishers were keen for me to write another. “How about one on water wars?” they asked. It seemed a good idea. The 1990s had seen cataclysmic forecasts, such as former World Bank vice-president Ismail Serageldin’s often-quoted 1995 prophecy that, although “the wars of this century were fought over oil, the wars of the next century will be fought over water”.

This and similar warnings entered the zeitgeist. Tony Allan, a social scientist at King’s College London and the School of Oriental and African Studies (SOAS) in London, summarized<sup>4</sup> the not-so-subtle argument as “if you run out of water you reach for a Kalashnikov or summon the air strike”.

I had no difficulty finding sources to back up this argument, and I set about writing chapters on the Jordan, the Nile and the Tigris–Euphrates river systems. My chapter choice relied on what seemed a perfectly reasonable assumption: that water scarcity was governed by the presence or absence of flowing water.

Allan had made the same assumption a few decades earlier when he set out to study the water situation in Libya. By the mid-1980s, water stress in North Africa and the Middle East had worsened; but Allan began to question his assumptions when he found no sign of the widely predicted water wars. Instead, the burgeoning populations of the Middle Eastern economies had no apparent difficulties in meeting their food and water needs. Allan had been forced to grapple with a situation in which people who are short of water do not necessarily fight over it.

### Invisible water

Allan’s earlier thinking about water wars began to change after meeting the late Gideon Fishelson, an agricultural economist at Tel Aviv University, Israel. Fishelson argued that it is foolish for Israel, a water-short country, to grow and then export products such as oranges and avocados, which require a lot of water to cultivate. Fishelson’s work prompted Allan to realize that water ‘embedded’ in traded products could be important in explaining the absence of conflict over water in the region.

As a global average, people typically drink one cubic metre of water each per year, and use 100 cubic metres per year for washing and cleaning. Each of us also accounts for 1,000 cubic metres per year to grow the food we eat. In temperate climates, the water needed to produce this food is generally taken for granted. In arid regions, Allan described how people depend on irrigation and imported food to fulfill these needs. Imported food, in particular, saves on the water required to cultivate crops.

The relationship of food trade to water sustainability is often not obvious, and often remains invisible: no political leader will gain any popularity by acknowledging that their country makes up the water budget only by importing food. Allan saw through this to document how the water budgets of the Middle East were accounted for without conflict.

Allan wrote about embedded water for a few years without it exciting any comment. Then, on a dark Monday afternoon in November 1992, during a routine SOAS seminar, somebody used the term ‘virtual’ water to describe the same concept. Allan realized this attention-grabbing word, in vogue with the computer-

literate younger generation, would catch on better than his own term. And he was right: “From there on it flew,” he says.

Allan’s work explained how, as poor countries diversify their economies, they turn away from agriculture and create wealth from industries that use less water. As a country becomes richer, it may require more water overall to sustain its booming population, but it can afford to import food to make up the shortfall<sup>5</sup>.

Areas seemingly desperate for water arrive at sustainable solutions thanks to the import of food, reducing the demand for water and giving an invisible boost to domestic supplies. Political leaders can threaten hostile action if their visible water supplies are threatened (a potentially useful political bluff), while not needing to wage war thanks to the benefits of trade.

### Sources of war

Israel ran out of water in the 1950s: it has not since then produced enough water to meet all of its needs, including food production. Jordan has been in the same situation since the 1960s; Egypt since the 1970s. Although it is true that these countries have fought wars with each other, they have not fought over water. Instead they all import grain. As Allan points out, more ‘virtual’ water flows into the Middle East each year embedded in grain than flows down the Nile to Egyptian farmers.

Perhaps the most often quoted example of a water war is the situation in the West Bank between Palestinians and Israel. But as Mark Zeitoun, senior lecturer in development studies at the University of East Anglia in Norwich, UK, has explained, contrary to what both the mass media and some academic literature say on the subject, while there is conflict and tension — as well as cooperation — there is no ‘water war’ here either<sup>6</sup>.

Ten million people now live between the Jordan River and the Mediterranean Sea. If they were to be self-sufficient in food, they would need ten billion cubic metres of water per year. As it is, they have only about one-third of that: enough to grow 15–20% of their food. They import the rest in the form of food. When it comes to water for domestic and industrial use, the rainfall and geology of the West Bank alone should provide enough water for the population there: Ramallah has a higher annual average rainfall than Berlin. But today, water for even these needs is scarce.

**“In five decades, there were no formal declarations of war over water.”**



ILLUSTRATION BY J. FIELD

Power struggles and politics have led to overt and institutionalized conflict over water — but no armed conflict, as there is over borders and statehood. Instead, Palestinian and Israeli water professionals interact on a Joint Water Committee, established by the Oslo-II Accords in 1995. It is not an equal partnership: Israel has de facto veto power on the committee. But they continue to meet, and issue official expressions of cooperation, even in the face of military action. Inequitable access to water resources is a result of the broader conflict and power dynamics: it does not itself cause war.

The Nile Basin Initiative, launched in 1999 and encompassing nine nations, is another example of the way in which wider geopolitical and economic factors help to balance water allocation. Historically, vast differences in the political clout of nations across which, or along which, a river flows have resulted in unequal water division. Under the 1959 Nile Waters Agreement between Egypt and Sudan, Egypt has had rights to 87% of the Nile's water, with Sudan having rights to the rest. Ethiopia, whose highlands supply 86% of Nile water, does not even figure in the agreement: continuing conflicts weakened the agreement to a point where Ethiopia has been unable to press a claim. But Egypt's desire to consolidate its economic development necessitates that it now come to better terms with its neighbours, improving prospects for local trade. So Egypt is willing to

engage in the multilateral initiative to cooperate more on matters such as hydroelectric power development, power-sharing cooperatives, river regulation and water-resources management.

Likewise, although India and Pakistan have fought three wars and frequently find themselves in eyeball-to-eyeball confrontation, the 1960 Indus Waters Treaty, arbitrated by the World Bank, has more than once helped to defuse tensions over water.

### Oil and water don't mix

Yet the myth of water wars persists. Climate change, we are told, will cause water shortages. The Intergovernmental Panel on Climate Change estimates that up to 2 billion people may be at risk from increasing water stress by the 2050s, and that this number could rise to 3.2 billion by the 2080s<sup>7</sup>.

Water management will need to adapt. But the mechanisms of trade, international agreements and economic development that currently ease water shortages will persist. Researchers, such as Aaron Wolf at Oregon State University, Corvallis, and Nils Petter Gleditsch at the International Peace Research Institute in Oslo, point out that predictions of armed conflict come from the media and from popular, non-peer-reviewed work.

There is something other than water for which shortages, or even the perceived threat

of future shortages, does cause war — oil. But the strategic significance of oil is immeasurably higher than that of water. Serious interruptions of oil supplies would stop highly developed economies in their tracks. Oil is necessary for a developed economy, and a developed economy provides for all the needs of its citizens, including water. People in developed economies do not die of thirst.

My encounter with Allan's work killed my book. I offered to revise its thesis, but my publishers pointed out that predicting an absence of war over water would not sell.

Book or no book, it is still important that the popular myth of water wars somehow be dispelled once and for all. This will not

only stop unsettling and incorrect predictions of international conflict over water. It will also discourage a certain public resignation that climate change will bring war, and focus attention instead on what politicians can do to avoid it: most importantly, improve the conditions of trade for developing countries to strengthen their economies. And it would help to convince water engineers and managers, who still tend to see water shortages in terms of local supply and demand, that the solutions to water scarcity and security lie outside the water sector in the water/food/trade/economic development nexus. It would be great if we could unclog our stream of thought about the misleading notions of 'water wars'.

**Wendy Barnaby** is editor of *People & Science*, the magazine published by the British Science Association.

e-mail: w.barnaby@btinternet.com

1. *Water in a Changing World* United Nations World Water Development Rep. 3; available online at [www.unesco-wwap.org/wwdr3/media-kit/documents/WWDR3%20low%20res.pdf](http://www.unesco-wwap.org/wwdr3/media-kit/documents/WWDR3%20low%20res.pdf) (2009).
2. Yoffe, S., Wolf, A. T. & Giordano, M. J. *Am. Wat. Resour. Assoc.* **39**, 1109–1126 (2003).
3. Barnaby, W. *The Plague Makers: The Secret World of Biological Warfare* (Continuum, 2002).
4. Allan, J. A. *Government and Opposition* **40**, 615–617 (2005).
5. Allan, A. J. *The Middle East Water Question* (I. B. Tauris, 2000).
6. Zeitoun, M. *Wat. Int.* **32**, 105–120 (2007).
7. *Climate Change and Water* Intergovernmental Panel on Climate Change; available online at [www.ipcc.ch/pdf/technical-papers/climate-change-water-en.pdf](http://www.ipcc.ch/pdf/technical-papers/climate-change-water-en.pdf) (2008).



## BOOKS &amp; ARTS

## When wise words are not enough

**Thomas Homer-Dixon** argues that opinion-makers must demonstrate a better grasp of how societies rise and fall if they are to steer nations successfully through many of this century's major crises.

These are bewildering times. One moment the global economy is booming and stock markets are soaring; the next, trillions of dollars of wealth have vanished, and we are on the cusp of a global depression. Oil prices rocket upwards as leading oilmen talk of worldwide shortages, then they plummet amid a worldwide glut. The United States seems to be in terminal decline, written off by pundits as a has-been hegemon, only to receive a new lease of life, many perceive, with the election of an extraordinary leader.

In a world reeling from surprise, where once-in-a-lifetime events seem to happen every month, two things seem to be constant. The first is the inadequacy of expertise. Although the people we have anointed as experts might not admit it, they are as bewildered by the world's turbulence as the rest of us. They are also little better at predicting what is going to happen next. The second constant is a pervasive feeling of insecurity. The things we assume to be bedrock truths around which we can organize our lives — scientific theory, moral precepts, political institutions or perhaps the timeless rhythms of nature — seem to be increasingly under assault.

In two books that offer erudite assessments of the dangers facing humankind this century, Vaclav Smil and Chris Patten address these matters in sharply different ways. In *Global Catastrophes and Trends*, Smil, a Canadian scientist of prodigious productivity and extraordinary disciplinary breadth, basically says "get used to it". Many of the vital natural and social systems around us are so complex that deep uncertainty characterizes their behaviour, and predicting this behaviour is near impossible. Thankfully, many of the threats to our wellbeing highlighted by the media are exaggerated — often wildly so. Although there are reasons for concern about where humankind is going, we need to remember that insecurity is part of the human condition. Catastrophe is too, but it is less likely than we imagine. Overall, given the admirable human capacity to adapt and change, the human prospect is far brighter than many assume.

**Global Catastrophes and Trends:**

**The Next Fifty Years**

by Vaclav Smil

MIT Press: 2008. 320 pp. \$29.95, £19.95

**What Next? Surviving the  
Twenty-First Century**

by Chris Patten

Allen Lane: 2008. 496 pp. £25



P. PARKS/AFP/GETTY IMAGES

The biggest threats to our existence are likely to be pandemics and wars, not asteroid hits or terrorism.

Patten, the last governor of Hong Kong, a cabinet minister in Margaret Thatcher's UK government and currently the co-chair of the International Crisis Group, a conflict-analysis agency, is less sanguine about the formidable challenges we face — including climate change, resource depletion, global criminality, widening gaps in wealth and possible nuclear terrorism. But in his book *What Next?*, he shares Smil's optimism about humankind's capacity to cope. With scientific practices and the reasoning brain that

evolution has given us, we can understand the basic contours of our challenges, if not predict exactly how they will shape our future. With freedom, democracy, the rule of law and strong, competent states — all legacies of the brutal

process of past European state building — we have the social and institutional architecture needed to find and implement solutions. These powerful assets allow us to choose between better and worse futures. The open question is whether we have the will to choose correctly.

Both books are grounded in the mastery of a staggering amount of knowledge — in Smil's case, largely technical and scientific; in Patten's, more social and political. Each is steeped in the wisdom that comes from

decades of experience and reflection. And each weaves argument and evidence around a clear structure of analysis.

Yet both books ultimately disappoint, for reasons that say less about the books themselves than about the largely unrecognized gravity of humankind's current predicament.

Smil's analysis is guided by a clear time horizon and a simple distinction. What, he asks, are the factors that could plausibly and fundamentally affect the human condition in the next five decades? He distinguishes between catastrophes and trends — between "fatal discontinuities" that could cause a sudden turn for the worse in the human prospect, and long-term incremental change that is barely noticed year-on-year but that could, over time, alter deep characteristics of our world.

Smil's discontinuities include large asteroid impacts, mega-eruptions of volcanoes, influenza pandemics, transformational wars and terrorist attacks. After assessing the best evidence available, he concludes that the risks of pandemics and major wars are large enough — approaching 100% for a war that will kill five million people sometime in the next 50 years — to justify significant investment in prevention. We should worry much less about asteroids and volcanoes, and current evidence does not justify our preoccupation with terrorism. He is indefensibly fatalistic about nuclear terrorism. Prevention, he says, cannot thwart all attacks, deterrence does not work against

zealots and we cannot effectively prepare for an attack in advance. But in reality, much can be done to prevent a nuclear strike in a major city; most importantly, nations can work harder to secure and ultimately destroy the world's stockpiles of highly enriched uranium.

As for unfolding trends, Smil includes in his list humankind's transition away from fossil fuels as a main energy source, shifting power relations among major nations, widening economic inequalities, climate change, perturbations of the global nitrogen cycle and antibiotic resistance. These trends do not sit together easily — for example, some implicate Earth's natural systems whereas others are mostly the product of human social and economic interactions — so Smil's analysis often comes across as a rather indigestible series of apples and oranges. Within the same chapter, for instance, he jumps from a treatment of the intrinsic power-density limits of renewable energy to an assessment of the geopolitics of world order.

Of greater concern is Smil's tendency to slip into polemic. His discussions of constraints on conventional oil supply and of climate change are selective and sometimes simply wrong. For example, he dismisses the possibility of abrupt climate shifts, based on a muddled interpretation of the science of millennial climate events recorded in ice cores (known as Dansgaard-Oeschger oscillations); he downplays recent research that strongly suggests sea levels will rise much faster than projected by the Intergovernmental Panel on Climate Change; and he wrongly suggests that the Gulf Stream is not driven partly by thermohaline circulation and that it does not contribute to Europe's warmth. He also repeats, as an illustration of how past predictions can be wildly wrong, the myth that scientists largely supported the early 1970s theory that global cooling was imminent.

Patten cannot claim Smil's technical expertise, but his book is more balanced. He acknowledges his biases and the limits of his knowledge, and gives fair consideration to contrary views. Some of his passages, especially those on China and India, are wonderfully rich. And his prescriptions are simultaneously deeply humane and pragmatically bounded by a long familiarity with the ways of the world — no mean feat.

Yet too often *What Next?* seems laboriously descriptive, absorbed in details of how we got to where we are, rather than focusing on what might happen later this century. The book's time horizon is only a few decades into the future, so its subtitle is misleading. And one is left with the disquieting sense that the state-based apparatus that Patten calls on to save humankind is not remotely up to the task. Patten is right to admire the power and

potential of Western institutions, but the challenges we now face, such as climate change, are decidedly different from those that such institutions evolved to address.

This brings us to why these books ultimately disappoint: in each case the analysis, although clearly structured, is not guided by an underlying theory of societal crisis. True, such a theory would not give the authors the power to predict the future with any precision, but it would help them to identify which combinations of factors might lead humankind over the precipice, and what we might do to avoid such a fate.

Recent research has highlighted fundamental causes of societal crisis. These include the convergence of interacting stresses that overloads the coping capacity of multiple components of society, especially the state; rising social and technical complexity that generates diminishing marginal societal returns; the

rapidly escalating cost of getting energy; high connectivity among people, institutions and technologies; and the declining redundancy and resilience of crucial systems.

One can make a credible argument, grounded in just the kind of evidence Smil and Patten use, that all of the above factors are now acting powerfully in humankind's tightly coupled, planetary, socio-ecological system. In the absence of a serviceable theory of societal crisis, in the end both books present little more than a laundry list of things we should worry about. For our children's and grandchildren's sake, we need much more, and fast.

**Thomas Homer-Dixon** is at the Balsillie School of International Affairs, Waterloo, Ontario N2L 6C2, Canada, and author of *The Upside of Down: Catastrophe, Creativity and the Renewal of Civilisation*.  
e-mail: tad@homerdixon.com

## Universe in a box

### Joseph Cornell and Astronomy: A Case for the Stars

by Kirsten Hoving

Princeton University Press: 2008.  
336 pp. \$49.50, £35

Sculptor Joseph Cornell made boxes. Intricate, three-dimensional montages of photographs, sky maps and paintings, neatly packed into wooden cases full of bric-à-brac: eggs, pipes, glasses, shells, stamps, chains and whatever objects fitted with his internal logic. And, even though he was considered one of the American exponents of surrealism, that epitome of irrationality, there was a logic. In a mesmerizing, if mildly flawed, attempt to immerse us in Cornell's crafted universes, art historian Kirsten Hoving uses the artist's fascination with astronomy to tease out the logic that underlies his work.

Cornell's life was complicated. Although he was from an affluent background, he had to support his younger brother who suffered from cerebral palsy, and until the late 1940s worked variously as a salesman, textile designer and in

a plant nursery. He was a staunch believer in the teachings of Mary Baker Eddy, the founder of Christian Science, and never forged a long-lasting relationship with a companion, spending most of his life in the same house in Utopia Parkway, a working-class neighbourhood of Flushing in Queens, New York. Yet, at the same time, his boxes and experimental films were admired in the high-octane art scene of New York, where he took part in the first surrealist exhibition at the Museum of Modern Art. Marcel Duchamp introduced him into the orbit of Peggy Guggenheim and her coterie, and from the early 1950s onwards, he was able to make a

living through his art.

Hoving does something ambitious and difficult: she identifies one important thread of his creative process and uses it to help us understand Cornell's art. Astronomy clearly played an important part in his work and life. His boxes, films and the countless dossiers that he stored at his house were littered with star maps, references to Albert Einstein and Arthur Stanley Eddington, solar eclipses and his fascination with space travel. Hoving uses these obsessions to relate a body of work that spans many decades. Albeit an effective choice, astronomy is only



Joseph Cornell's work mirrors the 1960s US cultural fascination with astronomy.

E. OWEN/ART RESOURCE/JOSEPH AND ROBERT CORNELL MEM. FOUND./DACS/VAGA



one thread of many, and it competes against his fascination with the natural sciences, art, poetry, modern dance and popular culture. Is one strand enough to understand him by?

The book's relentless focus on astronomy is to the detriment of a more comprehensive view of Cornell as an artist and as a man. Very little is said about his personal life, with only fleeting references to his reclusiveness, his close relationship with his brother and mother and his romantic trysts in later life. To compensate, Hoving extends her prose, clogging up the flow of the narrative. She feels obliged to describe the minutiae of individual artworks, to pepper the text with vignettes on cosmology and astronomy, and to make the case for a few of

the more tenuous links between his work and astronomy. To be fair, Hoving's arguments are on the whole convincing and at times enthralling, and her access to the debris of Cornell's life puts her in a unique position to back up many of her claims. But her thesis could have been made in half the number of pages. This is certainly a case in which less would have been more.

Yet, with its high-quality production and beautiful and wide-ranging illustrations, the book is extremely absorbing. I see it as Hoving's attempt to construct her own box, a carefully crafted piece laden with excerpts and images that penetrate Cornell's world and his obsessions. Furthermore, it is a showcase for the enthusiasm with the modern that

pervaded American popular culture during the twentieth century. By including newspaper articles and advertisements of the time in which astronomy and the space race were prominent, Hoving shows that Cornell was mirroring what he saw around him.

At a time when the interplay of science and art is ever more present in cultural life, and we begin to ask ourselves whether much of what is done is any good, *Joseph Cornell and Astronomy* gives us a good example of what works best: quiet fascination and obsession allied with genius. ■

**Pedro Ferreira** is a professor in astrophysics at the University of Oxford, Keble Road, Oxford OX1 3RH, UK, and author of *The State of the Universe*. e-mail: p.ferreira1@physics.ox.ac.uk

## A romp through science fiction

### Nanovision: Engineering the Future

by Colin Milburn

Duke University Press: 2008.  
296 pp. \$22.95, £17.99

Nanotechnology, we are often told, will change every aspect of our lives. It is easy to hype such claims because of the long timescales for approving a new technology for commercialization. The science itself can sound like science fiction at times, and fears about how nanotechnology will be realized in the future crop up in works of fiction. In *Nanovision*, historian of science Colin Milburn proposes that such speculation is an integral part of nanotechnology.

Milburn examines nanotechnology in a wealth of science-fiction texts, investigating such themes as the manipulation of matter, artificial life forms and the spread of nanobots as a 'grey goo' that ultimately consumes Earth. Many writers portray the science in a negative light, as the enabler of humankind's destruction or enslavement. Other imaginative works simply exploit differences between the physical laws that rule the nanoworld and the macroworld we live in, such as the effect of surface tension on a microscopic fictional organism. Milburn also looks at individuals who have straddled the fields of science and sci-fi, including molecular engineer and futurologist Eric Drexler, and carbon-nanotube discoverer Richard Smalley, who envisioned an elevator made of nanotubes stretching from Earth into space.

The book fills gaps that most nanoscientists have in their knowledge about the history of a field that has evolved so fast as to effectively obscure its founders. However, Milburn tries

too hard to prove the mix of science and fiction. In focusing on nanotechnology, he neglects the fact that forward thinking, technology forecasting and hyped claims are common to many scientific disciplines, such as genetics, energy research and space flight. More worryingly, he confuses theoretical science and fiction, missing the point that theories are not just speculation but are underpinned by mathematics, physical laws and reproducible calculations. His analysis also takes too literally the timetables and roadmaps contained in grant proposals and the like; documents that scientists consider to be indicative rather than prescriptive.

Some parts of the book, even the introduction, are difficult to follow. The writing is shrouded in the complex language of post-modern literary theory, and is often as dense as a hard-core physics paper. Milburn's references to gendered themes of male domination, even when discussing scientific instruments such as the scanning tunnelling microscope, are

sometimes involuntarily comical. Yet he ignores the large number of women in nanoscience and their views: no female nanoscientists are cited in the book.

*Nanovision* is not the place to seek clues about where nanotechnology is going, as it contains few scientific developments past the start of this decade. Beyond the hype, many useful applications — ranging from self-cleaning surfaces to sunscreen lotion containing nanoparticles of titanium dioxide, to nanocomposite materials in the car industry — are already on the market. Furthermore, in contrast to fears that the world will be devoured by grey goo, studies have shown that public perceptions of nanotechnology are mostly positive. Although there are real and tangible risks regarding the toxicity of nanomaterials, the scaremongering that fascinates Milburn is not such a pressing concern. But if you like a romp through themes that mix current nanoscience and literature in interesting ways, Milburn's book is a valuable read. Perhaps reality is stranger than fiction. ■

**Stefano Tonzani** is materials science editor at *Nature*.



Nanotechnology is often portrayed in fiction as destructive, but public perceptions are more positive.

# Disturbing the Arkansas river

## Over The River, A Work in Progress

Fondation de l'Hermitage, Lausanne, Switzerland

Until 24 May 2009.

The latest idea of the artists famous for wrapping Paris's Pont Neuf and Berlin's Reichstag in plastic is to cover parts of a 64-kilometre stretch of the Arkansas River in Colorado in aluminium-coated polypropylene. Christo sketched his first preparatory drawing for *Over the River* in 1992, but it won't come into being until the summer of 2012 at the earliest; hence the addendum to the title of this travelling exhibition, showing in Lausanne, Switzerland, until 24 May.

Although Christo does the drawings, he and his wife Jeanne-Claude are full artistic partners in every one of their "gentle disturbances". This is how they refer to their projects, which temporarily alter the relationship between sky and earth. *Over the River* will be in place in Colorado for just two weeks, after which it will be dismantled and recycled. All that will remain will be a few steel anchorage points, buried deep in the rock beneath the artificially raised banks that support the railway on the north side of the river and US Route 50 on the south.

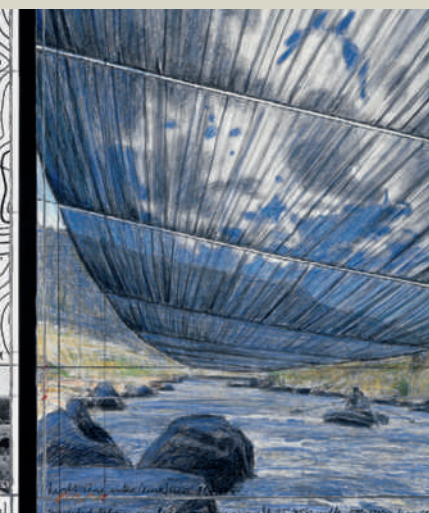
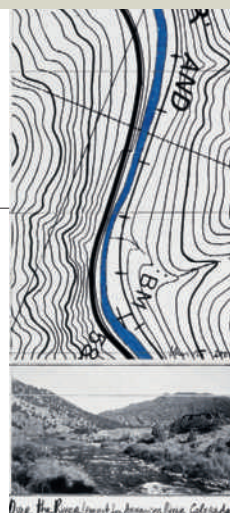
The canopy, consisting of around 900 panels and nearly 84,000 square metres of fabric, will

be broken up in places to make way for bridges, trees, rocks and other aesthetic interruptions. It will be translucent, so that people floating beneath it in rafts, canoes and kayaks can look up and see how the sky is changing. And when the wind blows, the artists say, it will undulate and shimmer "like a second, aerial river".

Christo and Jeanne-Claude chose the Arkansas River in 1996, having spent four years, on and off, reconnoitring in the Rocky Mountains. Full-sized and wind-tunnel tests followed, as well as an assessment of how the project will affect the environment — nothing like it has been done before. How will it withstand the temperatures in excess of 35 °C that are regularly recorded in Colorado in July? How will the mountain fauna respond to this disturbance?

A selection of Christo's preparatory drawings, as well as collages (pictured) of maps, photographs, pencil sketches and mock-up fabric panels, form the substance of the exhibition, along with black-and-white photos of people looking thoughtful on river banks. From these, Jeanne-Claude stands out with her frizzy hair and enormous hat, reminiscent of a beekeeper's.

As impressive as the project itself are the hoops the pair have jumped through to try to get permission to bring it off. This is chronicled



in a film diary that Wolfram Hissen started in 1994. The artists patiently put their case to bureaucrats and local residents who mostly seem to want to understand and to help; yet in an echo of creationists' objections to Charles Darwin's theories, some voice their anger at what they see as a couple of blasphemous egomaniacs bent on improving "God's work".

In the film, Christo explains that, having lived in Bulgaria when it was a Communist state, he has an urge to create art that has no ulterior motive. He and his wife have therefore devoted their lives to championing expensive, time-consuming and ephemeral projects. It may be their last — they are both more than 70 years old — and it may never happen. The locals will have the last word, and Hissen's film leaves you in doubt as to what that word will be. If *Over the River* ever sees the light of day, it will have emerged from its own chrysalis — an even bigger wrapping made entirely of red tape.

■  
Laura Spinney is author of the novel *The Quick*.

© CHRISTO 2001

# Not one to watch

## Watchmen

Film directed by Zack Snyder

In UK and US cinemas now.

Owing to complicated, interlocking storylines coupled with a gruesome take on humanity, Alan Moore's comic-book novel *Watchmen* was long considered to be unfilmable. Director Zack Snyder has at last made a film version. But the title should have stayed on the printed page.

*Watchmen* depicts the rise and fall of superheroes in an alternative 1985 in which the United States won the war in Vietnam, Richard Nixon is in permanent rule, the world is careening towards nuclear war, and a terrible lab accident created Doctor Manhattan, a super-being capable of manipulating the quantum universe.

The depth of character and story development in the 12-issue series, published by DC Comics in 1986–87, transformed the genre. The books were brutal, dystopian and had a deeply pessimistic view of the human condition. *Watchmen*



Quantum morality: Dr Manhattan in *Watchmen*.

grappled with intractable moral issues, many arising from cold-war apocalyptic nightmares. It asked if humanity would destroy itself, if that self-destruction could be prevented, and at what cost — and it questioned the role of science in preservation versus destruction.

Some elements of *Watchmen*'s failure as a film were predestined. It sinks beneath the weight of backstory — necessary for the uninitiated but yawningly dull. The most colossal blunder is the

casting of Matthew Goode as the smartest man in the world, Adrian Veidt, also known as the superhero Ozymandias. Veidt, who propels the story to its awful conclusion, should be magnetic but instead comes across as repulsively foppish. It is impossible to accept him as the executor of the most diabolical practical joke ever.

The end of the film fails, especially in terms of science ethics. In the book, saving the world from nuclear holocaust is accomplished through abhorrent yet inspired acts of technological and artistic creation. The result shocks — a Machiavellian decision with cataclysmic consequences, leaving the reader to dine on a stew of ambiguity. Was this murderous apotheosis of science ultimately required for the preservation of humanity? What is the cost to society of Veidt's salvation plan, or even of modern efforts such as the International Space Station? In the film, the creative act is replaced by one of theft and dissimulation, and the choice for the viewer is stripped of subtlety.

The book and film are summarized by asking "Who watches the Watchmen?" For the film, the answer should be, not many. Buy the book. ■

Michael White is an associate editor at *Nature*.

PARAMOUNT PICTURES



## SPECTROSCOPY

# Handedness in quick time

Patrick H. Vaccaro

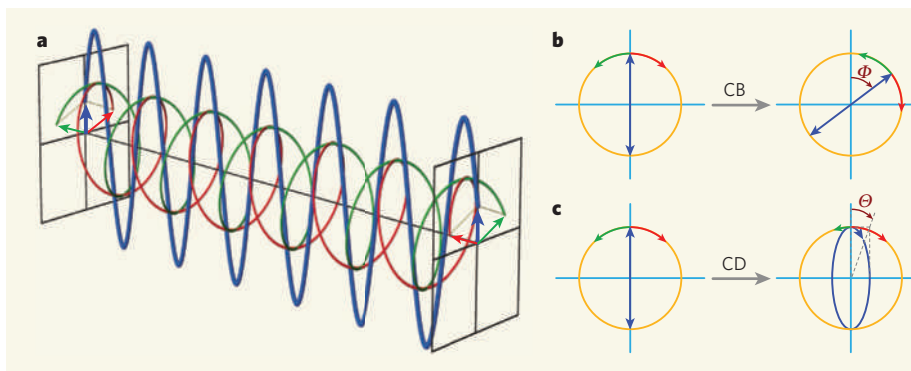
**The handedness of chiral molecules can be probed spectroscopically, but acquiring data can take hours, which is a problem for time-resolved studies. The latest method records such data in a flash.**

Most daily encounters with chirality and chiral recognition go unnoticed, including such mundane events as putting on shoes, shaking hands with someone and admiring the helical patterns of seashells. But the concept that certain objects and interactions have an intrinsic 'handedness' permeates the entire fabric of science, and has a crucial role in diverse physical, chemical and biological processes. Of special importance for the molecular sciences are the different chemical and biochemical reactivities displayed by enantiomeric (mirror-image) compounds. Indeed, many drugs are single enantiomers; their mirror-image versions are often ineffective, or even harmful.

The development of tools to discriminate between enantiomers and to work out how chiral molecules interact is therefore an ongoing challenge. In particular, the unique signatures that can be obtained using optical probes that are based on chiral — circularly polarized — light (Fig. 1a) represent a long-standing, yet constantly evolving avenue of research<sup>1</sup>. On page 310 of this issue, Rhee *et al.*<sup>2</sup> demonstrate a significant advance that promises to open new vistas in the realm of chiroptical spectroscopy, including the tantalizing possibility of interrogating chiro-specific phenomena with femtosecond temporal resolution (1 femtosecond is  $10^{-15}$  seconds).

When electromagnetic radiation passes through a randomly oriented (isotropic) ensemble of non-chiral molecules — such as those in a gas, liquid or glass — its state of polarization, defined by the direction of oscillation of the electric-field vector, remains unaltered. This implies that the sample's frequency-dependent refractive index,  $n$ , and absorption index,  $\kappa$ , are independent of optical polarization. These parameters govern the speed and intensity of a wave propagating through the medium.

In contrast, light traversing a chiral medium experiences an intrinsic helical anisotropy, commonly referred to as optical activity, which causes right-circularly ( $R$ ) and left-circularly ( $L$ ) polarized light to undergo distinct retardation ( $n_L \neq n_R$ ) and attenuation ( $\kappa_L \neq \kappa_R$ ) processes<sup>1</sup>. The differential retardation ( $\Delta n = n_L - n_R$ ) leads to the phenomenon



**Figure 1 | Optical polarization and optical activity.** **a**, The oscillating electric-field vector of linearly polarized electromagnetic radiation (blue) can be broken down into equal contributions of right-circular (red) and left-circular (green) polarized waves. **b**, Chiral media are optically active — they impose an intrinsic anisotropy on linearly polarized light. Here, the electric-field vector oscillates in a plane defined by the blue arrow. The phenomenon known as circular birefringence (CB) causes one component of circular polarization to propagate faster than the other, so that the plane of the electric-field vector rotates by an angle  $\Phi$ . **c**, In circular dichroism (CD), one component of circular polarization is absorbed more strongly than the other, leading to net polarization ellipticity at an angle  $\Theta$ . Rhee *et al.*<sup>2</sup> report a technique for acquiring information about the optical activity of chiral compounds on a femtosecond timescale, in principle opening the way to ultrafast time-resolved studies of chiro-specific chemical and biochemical processes.

of circular birefringence (CB, Fig. 1b), the observable effect of which is the rotation of linearly polarized light from its original angle; the differential attenuation ( $\Delta\kappa = \kappa_L - \kappa_R$ ) causes circular dichroism (CD, Fig. 1c), in which  $R$ - and  $L$ -polarized light are absorbed unequally.

The two enantiomeric forms of a chiral compound display optical activities ( $\Delta n$  and  $\Delta\kappa$ ) of equal magnitude, yet opposite sign, reflecting the chiro-specific nature of their interactions with oscillating electric and magnetic fields. The patterns obtained by recording CB or CD as a function of wavelength thus afford a spectral 'fingerprint' for a given enantiomer and its environment. The majority of such measurements rely on electronic excitations caused by radiation in the visible and ultraviolet regions of the spectrum. But analogous methods involving infrared light — which excites molecular vibrations — have also emerged as powerful probes of molecular structure and function<sup>1,3</sup>. Rhee *et al.* build on the techniques of vibrational circular dichroism (VCD)<sup>4,5</sup>, in which the unique chiroptical behaviour displayed by each vibrational feature of an

infrared absorption spectrum conveys information about the configuration and conformation of the molecule under investigation.

Unfortunately, the chiroptical response of an isotropic solution is exceedingly small, and so any viable probe of CD must be able to distinguish a minuscule chiral absorption signal from a substantially larger background of achiral absorption. For VCD (where the ratio of signal to background absorption is typically  $10^{-5}$ – $10^{-6}$ ), exquisitely sensitive detection schemes are needed to isolate the slight differences in sample absorption revealed by alternating bursts of  $R$ - and  $L$ -polarized infrared light. Even when the most efficient data-collection methods are used (such as Fourier-transform infrared spectroscopy<sup>4</sup>), hours of continuous averaging are often required to produce statistically meaningful VCD spectra, thus limiting the usefulness of the technique for time-resolved studies.

But Rhee *et al.*<sup>2</sup> have developed a strategy that permits both vibrational CB (known as vibrational optical rotatory dispersion, VORD) and VCD to be recorded without the need to

differentially discriminate minute signals. In their technique, a short infrared pulse is used to coherently excite vibrations in a dissolved chiral compound, thereby creating a macroscopic assembly of oscillating dipoles. This induced 'polarization' radiates the vibrational properties of the target molecules in the form of a signal known as a free-induction decay. This signal has chiro-specific information encoded in the amplitude and phase of its electric field — in the same way that the characteristic sound of a ringing bell reflects its size and shape.

So how is the weak chiroptical signal discriminated from the strong achiral background? The authors rely on the fact that linearly polarized light always develops another polarized component (perpendicular to the original direction of polarization) as it crosses a chiral medium (Fig. 1b, c). Using polarization analysers of exceptional quality aligned to detect only signals orthogonal to the impinging (linearly polarized) infrared beam, the authors were able to isolate the desired optical-activity response in their experiments.

To amplify and interpret the responses, Rhee *et al.* exploited the phenomenon of interference, in which the superposition of oscillating waves gives rise to distinctive patterns of new waves. They thus combined the free-induction decay from the chiral sample with an infrared reference pulse of much greater intensity. The coherently beating portion of the resultant pattern — the product of the weak signal field and the strong reference field — can be isolated and analysed using Fourier-transform manipulations<sup>6</sup>, yielding information about the sample.

Ultimately, the authors were able to ascertain a property of chiral molecules known as the complex susceptibility. This quantity consists of a part that is related to  $\Delta n$  and another part that is related to  $\Delta \kappa$ . In this way, VORD and VCD spectra can be acquired for each incident pulse of infrared radiation, with the timescale of each measurement essentially being given by the pulse duration (less than 100 femtoseconds).

Rhee *et al.* validated their approach by examining the enantiomers of a simple chiral molecule (limonene) dissolved in solution. They showed that the resulting spectra match those obtained by conventional VCD spectrometers and agree with theoretical VCD predictions. In principle, the techniques and concepts introduced in this work<sup>2</sup> should be applicable to a variety of chiroptical studies on systems in equilibrium, including those based on electronic (rather than vibrational) phenomena, provided that polarization analysers of sufficient quality are available for the targeted spectral region.

A time-resolved analysis of optical activity has yet to be performed, but the present results establish the basic principles required for studies in which changes in chirality induced by an abrupt pulse of 'pump' light can be monitored at subsequent points in time using a synchronously delayed infrared 'probe'. The resulting

arrays of VORD and VCD spectra should allow an unprecedented glimpse of the ultrafast structural and conformational changes that accompany chiro-specific chemical and biochemical transformations, such as those that occur during asymmetric catalysis or protein folding. Once the practical details are sorted out, the successful execution of time-resolved measurements will revolutionize our ability to interrogate chiral molecules and their interactions, in much the same way that the advent of high-speed photography overcame the limitations of visual perception. ■

Patrick H. Vaccaro is in the Department of Chemistry, Yale University, 225 Prospect Street, New Haven, Connecticut 06520-8107, USA. e-mail: patrick.vaccaro@yale.edu

1. Barron, L. D. *Molecular Light Scattering and Optical Activity* 2nd edn (Cambridge Univ. Press, 2004).
2. Rhee, H. *et al.* *Nature* **458**, 310–313 (2009).
3. Barron, L. D. *et al.* *Mol. Phys.* **102**, 731–744 (2004).
4. Freedman, T. B., Cao, X., Dukor, R. K. & Nafie, L. A. *Chirality* **15**, 743–758 (2003).
5. Stephens, P. J., Devlin, F. J. & Pan, J.-J. *Chirality* **20**, 643–663 (2008).
6. Lepetit, L., Chériaux, G. & Joffe, M. *J. Opt. Soc. Am. B* **12**, 2467–2474 (1995).

## CANCER

# The nuances of therapy

Lee M. Ellis and David A. Reardon

**Oncologists use drugs that limit a tumour's blood supply to prevent its growth. Although the initial effects of these drugs are beneficial to patients, new data suggest that their long-term effects warrant further study.**

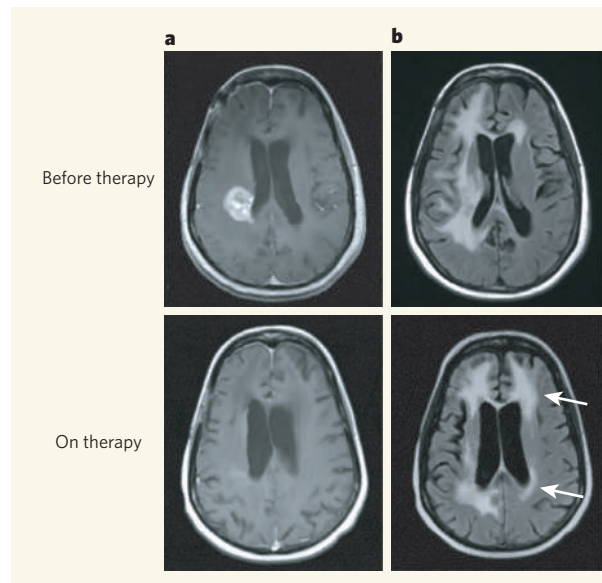
Angiogenesis — the formation of new blood vessels — is a hallmark of cancer, and allows tumour growth. Anti-angiogenic therapy offers great promise and is often used to treat cancer, either alone or in combination with chemotherapy. But, like all anticancer therapies, agents that inhibit tumour angiogenesis are prone to either intrinsic or acquired resistance. Pàez-Ribes *et al.*<sup>1</sup> and Ebos *et al.*<sup>2</sup> show in two preclinical (*in vitro* and animal) studies<sup>1,2</sup> published in *Cancer Cell* that, depending on treatment conditions, anti-angiogenic therapy could theoretically increase the likelihood of tumour invasiveness and spread.

One protein with a central role in promoting angiogenesis is vascular endothelial growth factor (VEGF), and anti-VEGF agents are therefore commonly used to treat cancer.

Nonetheless, as highlighted by several studies, host and tumour responses to loss of VEGF-mediated signalling can be complex<sup>3</sup>.

In mice, deletion of a single copy of the *VEGF* gene causes embryonic death<sup>4</sup>, suggesting that it is essential for survival. Loss of this gene specifically in endothelial cells, which line blood vessels, increases the probability of angiogenesis-related disease conditions such as thrombosis, haemorrhage and fibrosis<sup>5</sup>. What's more, VEGF inhibition affects not only tumour vasculature but also healthy host tissues. When VEGF activity is impaired in tumour-free mice, for example, compensatory pathways are activated that in theory could augment tumour invasion and metastasis in patients with cancer. Ebos *et al.* have previously shown<sup>6</sup> that anti-VEGF therapy increases levels of

**Figure 1 | The MRI evidence.** **a**, In agreement with the latest preclinical data<sup>1,2</sup>, MRI scans from a patient with recurrent glioblastoma show that, after treatment with the VEGF-neutralizing antibody bevacizumab and the chemotherapeutic agent irinotecan, the macroscopic 'enhancing' tumour disappears, consistent with a complete response. **b**, However, microscopic tumour infiltration to other brain regions (arrows) is detectable in the patient after this therapy, using a different type of MRI that highlights brain inflammation and swelling<sup>11</sup>.



D. A. REARDON



several blood-plasma proteins, including G-CSF, PlGF and SDF1, that may support invasion and metastasis by tumour cells. Response to oxygen shortage may also lead to a compensatory programme mediated by the mobilization of cells derived from the bone marrow, increasing tumour invasiveness and angiogenesis<sup>7</sup>.

The practical issue that Pàez-Ribes *et al.*<sup>1</sup> and Ebos and colleagues<sup>2</sup> now raise through pre-clinical studies is whether anti-VEGF therapy could shorten patients' long-term survival. Pàez-Ribes *et al.* studied the effects of either anti-angiogenic drugs or genetic ablation of the gene encoding a VEGF protein in mouse models of two cancers: glioblastoma, the deadliest of brain cancers, and pancreatic neuroendocrine tumours. They find that, although these treatments have antitumour effects, loss of VEGF signalling also increases tumour invasiveness and metastasis.

Ebos *et al.*<sup>2</sup> reached similar conclusions when they treated mice with VEGF-receptor inhibitors. Intriguingly, they found that whether these drugs prevented tumour growth or increased the likelihood of metastasis depended on tumour stage and the duration of treatment. Whereas sustained use of the drugs on established tumours significantly inhibited tumour growth, short-term treatment before the implantation of tumour cells into mice 'pre-conditioned' the animals for accelerated metastasis.

How do these findings relate to previous clinical observations? From experience in the clinic, single-agent anti-VEGF therapy is ineffective for treating non-small-cell lung cancer, and metastatic breast and colorectal cancers. However, anti-VEGF agents — specifically, bevacizumab, a VEGF-neutralizing antibody — improve the effects of chemotherapy in patients with these tumour types<sup>3</sup>. It is likely that such effects of combined therapy are due to complex interactions between anti-VEGF and chemotherapeutic agents, and so are not directly comparable with the preclinical data<sup>1,2</sup>.

In addition, a study<sup>8</sup> that analysed data from a prospective database of more than 1,900 patients with metastatic colorectal cancer indicated that the overall survival of those given bevacizumab throughout their treatment, in combination with various chemotherapeutic agents, was much better than the results of any clinical trial in which anti-VEGF therapy was stopped after initial disease progression — tumour growth despite therapy. This observation, combined with the data alluded to above, suggests that the preclinical findings of Pàez-Ribes *et al.* and Ebos *et al.* should not dissuade oncologists from using anti-VEGF therapy in combination with chemotherapy. However, these studies<sup>1,2</sup> may be more relevant to patients with a form of liver cancer called hepatocellular carcinoma, and possibly those with kidney cancer<sup>9,10</sup>. Single-agent anti-VEGF therapy in both of these patient groups improves survival

with no clinically evident deleterious effects.

Investigators are currently awaiting the results of several clinical trials in patients at risk of harbouring undetectable metastatic tumours and who receive chemotherapy together with bevacizumab in the adjuvant setting, in an attempt to prevent these 'micrometastases' from growing. In particular, the outcome of a trial in which patients with colorectal carcinoma received six months of combination therapy followed by six months of bevacizumab alone is expected soon. Results gathered from such studies should be evaluated not only for cancer-progression-free and overall survival rates, but also for the pattern of tumour recurrence — that is, whether patients with cancer recurrence demonstrate widespread metastases, as seen in the preclinical studies<sup>1,2</sup>.

Observations in patients with recurrent glioblastoma who were treated with anti-VEGF therapy with or without the chemotherapeutic agent irinotecan seem to support some of the findings of Pàez-Ribes *et al.* and Ebos and colleagues. Unlike historical salvage therapies, which achieve low response rates of nominal duration, most patients with recurrent glioblastoma respond to anti-VEGF therapy, as is evidenced by both improved MRI (magnetic resonance imaging) scans and improvement in their neurological deficits. What's more, in many, these gratifying responses are meaningfully durable<sup>11</sup>. Nevertheless, most patients become resistant to anti-VEGF therapy, showing marked infiltration of tumour cells, and with some tumours spreading to other regions of the brain, such as the contralateral hemisphere, brainstem and leptomeninges (Fig. 1); progression is also often quick and rapidly fatal<sup>11</sup>.

This observation supports the findings of Pàez-Ribes *et al.*<sup>1</sup> and Ebos *et al.*<sup>2</sup>, and together they indicate that, although anti-VEGF therapy may meaningfully improve the therapeutic outcome for some patients, resistance to this therapy can develop and may lead to more aggressive and widespread disease. These data therefore highlight the point that oncologists must carefully scrutinize clinical data, not only focusing on short-term endpoints, such as response rates and time to progression, but also patterns of failure, overall survival and quality of life.

On the basis of these findings<sup>1,2,11</sup>, should anti-VEGF therapy be limited in duration? As noted in the preclinical setting<sup>1,2</sup>, even a short duration of anti-VEGF therapy can increase the likelihood of invasion and metastasis by tumour cells; yet in the clinic, acceleration of tumour growth and/or metastasis is rarely (if ever) noted after initiation of therapy. Some investigators have noted a 'tumour flare' on cessation of anti-VEGF therapy, but these observations are anecdotal. One approach to resolving this dilemma is to identify increases in the levels of soluble markers or mediators of tumour growth and/or metastasis just before starting anti-VEGF therapy. Indeed, an increase in levels of PlGF and HGF — potential



## 50 YEARS AGO

*Television and the Child: An Empirical Study of the Effect of Television on the Young.* By Hilde T. Himmelweit, A. N. Oppenheim and Pamela Vince — The widespread interest aroused by this report is easily understood. It is the first study of its kind to be conducted among English-speaking peoples and, indeed, probably in the world ... In general, television does not cause listlessness, loss of sleep, bad dreams, lack of concentration, or eye strain. Children are not made more passive or become more aggressive, or kept away from youth clubs by television. Against this are a number of conclusions which, on balance, suggest that television, as at present administered, has more harmful than beneficial influence on children ... The main recommendation made by Dr. Himmelweit and her colleagues is that the B.B.C. and I.T.A. must collaborate much more over the balance of programmes.  
From *Nature* 21 March 1959.

## 100 YEARS AGO

In a letter to the *Times* of March 15, Prof. Osler directs attention to the useful work which is being done by the Italian Society for the Study of Malaria, founded ten years ago, for the prevention of malarial diseases. The society has promoted legislation for the gratuitous distribution of quinine, has prepared quinine in its most agreeable forms, and has introduced into practice the mechanical measures based on the defence of the habitation and the individual from the bites of mosquitoes. The result is that the mortality from malaria in Italy has declined from 16,000 in 1902 to about 4000 in 1908 ... the growth of our knowledge of the causation and prevention of malaria illustrates the stages through which so many of the great discoveries in medicine have had to pass, and is a striking example of the value of experimental methods in medical research.  
From *Nature* 18 March 1909.

50 & 100 YEARS AGO

mediators of resistance to anti-VEGF therapy — just before the resumption of growth of colorectal tumours and their metastases has been reported for this approach<sup>12</sup> in patients receiving chemotherapy plus bevacizumab.

Finally, although it was initially hypothesized that anti-angiogenic therapy would induce tumour dormancy, this is clearly not the case in the clinic. Thus, following an initial favourable response to chemotherapy combined with anti-VEGF therapy, single-agent anti-VEGF therapy should not be used until there are data from clinical trials to either support or refute this approach. ■

Lee M. Ellis is in the Departments of Cancer Biology and Surgical Oncology, The University of Texas MD Anderson Cancer Center, Houston,

Texas 77030, USA. David A. Reardon is at the Preston Robert Tisch Brain Tumour Center, Duke University Medical Center, Durham, North Carolina 27710, USA.

e-mail: lellis@mdanderson.org

1. Páez-Ribes, M. *et al.* *Cancer Cell* **15**, 220–231 (2009).
2. Ebos, J. M. L. *et al.* *Cancer Cell* **15**, 232–239 (2009).
3. Ellis, L. M. & Hicklin, D. J. *Nature Rev. Cancer* **8**, 579–591 (2008).
4. Ferrara, N. *et al.* *Nature* **380**, 439–442 (1996).
5. Lee, S. *et al.* *Cell* **130**, 691–703 (2007).
6. Ebos, J. M. L. *et al.* *Proc. Natl Acad. Sci. USA* **104**, 17069–17074 (2007).
7. Du, R. *et al.* *Cancer Cell* **13**, 206–220 (2008).
8. Grothey, A. *et al.* *J. Clin. Oncol.* **26**, 5326–5334 (2008).
9. Escudier, B. *et al.* *N. Engl. J. Med.* **356**, 125–134 (2007).
10. Llovet, J. M. *et al.* *N. Engl. J. Med.* **359**, 378–390 (2008).
11. Vredenburg, J. J. *et al.* *J. Clin. Oncol.* **25**, 4722–4729 (2007).
12. Kopetz, S. *et al.* *Proc. 2009 Gastrointest. Cancers Symp.* Abstr. 292 (Am. Soc. Clin. Oncol., 2009).

## GEOPHYSICS

# Hot blanket in Earth's deep crust

Jean Braun

**Studies of rocks from Earth's crust suggest that the lower crust is a good thermal insulator. The knock-on effects of this finding are many — one being the crust's increased potential to generate more magma.**

We know little about the temperature of the deepest parts of Earth's continental crust, which can be 30–40 kilometres below the surface. We cannot measure temperature directly beyond a few kilometres down, at the bottom of deep mines or drill holes. Yet temperature governs many geological processes, including the generation of magmas through the melting of rocks in the lower crust and the mechanism by which rocks deform. In the crust, heat from the underlying mantle is mainly transported by conduction. The efficiency of the process is regulated by the crust's thermal conductivity: a high conductivity results in efficient heat flow and a relatively low temperature at the base of the crust; conversely, a low conductivity results in poor conduction of heat and a high temperature at the base of the crust — the lower crust is a good insulator and thus acts as a blanket over the mantle.

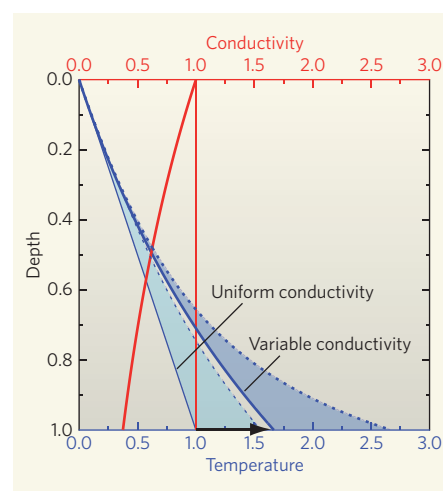
Like most materials, rocks become less efficient at transporting heat (their conductivity decreases) as temperature increases, especially in the vicinity of their melting temperature. But measuring thermal conductivity at high temperatures is difficult because, close to the melting temperature, heat is transported not only by conduction but also by radiation. This implies that most existing high-temperature rock-conductivity measurements have been overestimated. Consequently, in many applications, rock conductivity is assumed to be constant throughout the crust. Whittington *et al.*<sup>1</sup> (page 319 of this issue) now provide more accurate data on a variety of crustal rock samples that demonstrate a strong dependence of rock

thermal conductivity on temperature. This means that the conductivity of rocks at temperatures existing in the lower crust is much lower — by as much as 50% — than previously recognized. The implication is that the lower crust may be hotter than we thought (Fig. 1).

By using a recently developed technique known as laser-flash analysis, which consists of subjecting one end of a small (10 mm × 1 mm) rock sample held at some temperature in a hot furnace to a laser pulse of heat, Whittington and colleagues were able to measure the time taken for the heat to reach the other end. They then used a carefully calibrated model of heat transport in the sample to extract accurate estimates of its thermal conductivity.

The authors' observation<sup>1</sup> that rock conductivity varies strongly with temperature, and the related realization that the lower crust has a lower conductivity and is thus warmer than we thought, has far-reaching implications for our understanding of fundamental magmatic and tectonic processes (Fig. 2). The conductivity of a rock is more correctly described by its thermal diffusivity — the ratio of conductivity to the product of the heat capacity (the ability to store heat) and density. It is this that controls the rate at which the crust responds to a perturbation in temperature such as is caused by the injection of magma from the underlying mantle or by 'strain heating' in shear zones (Fig. 1).

Reduced conductivity in the lower crust would thus prevent the dissipation of the heat brought into the crust by intrusion of basaltic rock from the mantle; temperatures within and



**Figure 1 | Thermal conductivity and temperature in Earth's crust.** The temperature profile of Earth's crust is shown for two scenarios of conductivity for a crust subjected to a constant basal heat flux: uniform rock conductivity (thin blue line) and variable rock conductivity (thick blue line). Whittington *et al.*<sup>1</sup> show that a non-uniform conductivity across the crust results in an increase in basal temperature, as is indicated by the arrow. The shaded areas show the difference in the crust's response to an instantaneous perturbation in the basal heat flux, such as that caused by a basaltic intrusion or shear heating: the light-blue area denotes the resulting change in temperature for a uniform conductivity; dark blue is for variable conductivity. In the latter case, the reduced conductivity of the lower crust decreases the upward dissipation of the thermal event and enhances heating of the lower crust. The red lines show the two conductivity profiles. The depth scale spans the entire crust (from the surface down to 35–40 km); the temperature scale varies between 0 °C and an upper range of 500–800 °C (this upper limit depends on the assumed contribution from heat production due to radioactivity).

around the intrusion should be higher than we currently predict, and we should expect more rapid and efficient melting of the lower crust for the same amount of heat injected from the mantle. This effect has important consequences for our understanding of arc or andean-type volcanism (that is, volcanism associated with subduction zones). Lower conductivity of magma also helps its ascent through the crust, because it slows down the cooling and 'freezing' that occur by contact with the surrounding, colder rocks.

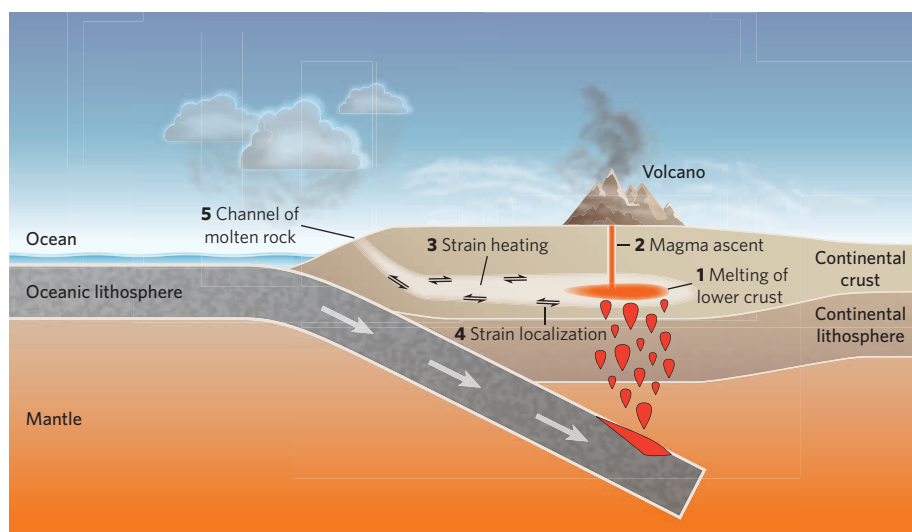
The deformation of the crust in response to tectonic forces is commonly accompanied by strain localization: that is, by the formation of narrow, ductile shear zones. During deformation at high stress levels, mechanical energy is transformed into heat (a process called strain heating), which is dissipated from the shear zone by conduction. A reduced thermal conductivity at high temperature would limit heat dissipation, resulting in much hotter shear zones than commonly assumed. As



Whittington *et al.*<sup>1</sup> demonstrate, this may even lead to melting inside the shear zone, causing further localization of the deformation within the shear zone. But this result should be interpreted with caution, as strain heating depends strongly on the evolution of stress within the zone, which itself depends on the assumed deformation mechanism and is currently the subject of much debate<sup>2</sup>. Enhanced strain heating could also explain the formation of larger than expected volumes of granitic magmas produced by partial melting of deep crustal rocks<sup>3</sup>. This process, referred to as anatexis, is observed in many mountain ranges, most notably in the granitic belts associated with the Main Central Thrust fault in the Himalayas.

A recent and vigorously debated model for the evolution of large, hot mountain belts, such as the Himalayan–Tibetan system<sup>4</sup>, relies on the formation of a ductile, partially molten lower crust that may be extruded at Earth's surface in regions of high precipitation and thus surface erosion. The high-temperature conditions necessary for partial melting of the lower crust are thought to be achieved by increased heat production due to radioactivity, resulting from a thickened crust and/or shear heating. A reduced conductivity in the lower crust would significantly reduce the time necessary for such radiogenic heating to take place, and would potentially lead to higher crustal temperatures. This hypothesis, and many others concerning the dynamics of mountain belts, should now be revisited in view of Whittington and colleagues' new measurements<sup>1</sup> of thermal conductivity.

Also in need of reassessment are most quantitative models of Earth's dynamical behaviour,



**Figure 2 | Implications of a lower crust that is warmer than expected.** Whittington and colleagues' observation<sup>1</sup> that the lower crust is a good thermal insulator and is thus warmer than previously recognized has many consequences for magmatic and tectonic processes in convergent plate settings such as subduction zones: **1**, more efficient melting and mixing of the lower crust following mantle-derived basaltic intrusion; **2**, more rapid and efficient ascent of the resulting magma through the crust; **3**, enhanced strain heating during tectonically driven deformation of the lower crust; **4**, enhanced strain localization; **5**, quicker development of a lower-crust channel of molten rock that may be extruded at Earth's surface in regions of high precipitation and thus surface erosion.

because heat transport is such an important process inside the Earth. For example, a reduced crustal conductivity would also imply a higher mean temperature in the underlying mantle, especially during the early, 'hotter' stages of the planet's evolution. This, in turn, has direct implications for our understanding of the early Earth's differentiation and the distribution of elements in its various reservoirs (the crust, mantle and core).

Jean Braun is at the Géosciences Rennes CNRS UMR 6118, Université de Rennes 1, Rennes F-35042, France.  
e-mail: jean.braun@univ-rennes1.fr

1. Whittington, A. G., Hofmeister, A. M. & Nabelek, P. I. *Nature* **458**, 319–321 (2009).
2. Regenauer-Lieb, K. *et al.* *Nature* **442**, 67–70 (2006).
3. Harrison, T. M., Grove, M., Lovera, O. M. & Catlos, E. J. *J. Geophys. Res.* **103**, 27017–27032 (1998).
4. Beaumont, C. *et al.* *Nature* **414**, 738–742 (2001).

## DINOSAURS

# Fuzzy origins for feathers

Lawrence M. Witmer

**Cretaceous fossil deposits in China are famous for their feathered dinosaurs. But the surprising discovery of a herbivorous dinosaur with a filamentous coat raises fresh questions about the evolution of feathers.**

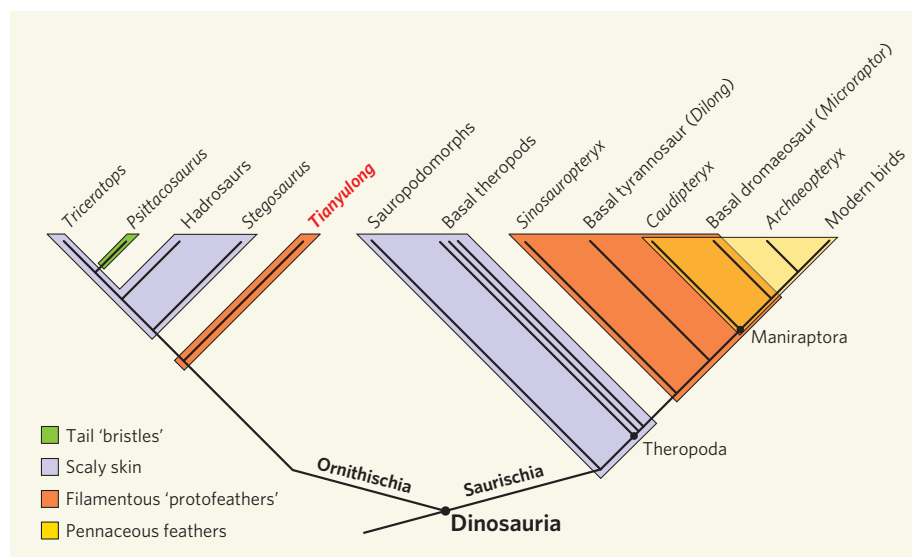
Liaoning Province in northeastern China is renowned for the fossils that document, in often vivid detail, virtually the entire biota that lived over a period of several million years during the Early Cretaceous (about 125 million years ago)<sup>1</sup>. Although exquisite fossils of diverse vertebrates, invertebrates and plants have been recovered, it's the spectacular feathered dinosaurs that have received most attention<sup>2–4</sup> and caused much controversy<sup>5,6</sup>. On page 333 of this issue, Zheng and colleagues<sup>7</sup> present the discovery of a small dinosaur, *Tianyulong confuciusi*, from the Yixian Formation of Liaoning, that promises to send the debate on dinosaur feathers in

a totally new direction — and a confusing direction, at that.

Even without preservation of portions of its skin, *Tianyulong* would be a notable find, because its genealogical ties are to a group of herbivorous dinosaurs, called heterodontosaurids, that had undergone their evolutionary radiation 70 million years earlier, making *Tianyulong* a 'living fossil' in its own time. Heterodontosaurids used to be regarded as a fairly obscure group, related to the more famous duck-billed hadrosaurs. But recently, heterodontosaurids have taken centre stage as the most evolutionarily basal branch of the entire great radiation of herbivorous dinosaurs, the

Ornithischia (Fig. 1, overleaf), that included not only hadrosaurs but also *Triceratops*, *Stegosaurus* and a host of related animals<sup>8</sup>. Were it not for its skin, *Tianyulong* would be important as a late-surviving twig of this branch of the ornithischian family tree, and the first of its kind known from Asia. But the fossils of *Tianyulong*, splayed on a stone slab, reveal three patches of long filaments reminiscent of structures thought to be the evolutionary progenitors of feathers. The only problem is that *Tianyulong* isn't supposed to have anything like feathers.

Before the 1990s, life was simple: feathers were thought to be an exclusively avian attribute, found in all birds today and extending back to the iconic *Archaeopteryx* in the Late Jurassic, some 150 million years ago. The discovery of very bird-like feathers, complete with shaft and vanes (pennaceous feathers), in some of the predatory theropod dinosaurs found in Liaoning (such as *Caudipteryx* and *Micro-raptor*) rocked the scientific world<sup>2,4,9</sup>, because the feathered dinosaurs were outside the evolutionary group of acknowledged birds. Still, it wasn't completely unexpected, in that these dinosaurs are representatives of the theropod



**Figure 1 | Dinosaur relationships and skin characteristics.** The dichotomy between feathered birds and scaly reptiles was demolished by the discovery of true pennaceous feathers in the non-avian maniraptoran dinosaurs thought to be closest to birds. More controversial have been the filamentous skin structures, variously regarded as external 'protofeathers' or internal structural fibres. Zheng and colleagues' discovery<sup>7</sup> of filamentous structures in *Tianyulong* further complicates the picture, in that this is an ornithischian dinosaur far removed from the ancestry of birds. Many other dinosaurs, such as other ornithischians and basal saurischians, had reptilian, scaly skin. So, were the ancestral dinosaurs fuzzy animals cloaked in 'protofeathers', which were subsequently lost multiple times in later groups? Or were dinosaurs primitively scaled, and did later groups independently evolve wispy, feather-like or even bristly skin coverings?

group, Maniraptora, that is skeletally the most bird-like, and from which birds are widely thought to have evolved. Feathers became just one more character showing that link, albeit a compelling one. But *Tianyulong* is not at all closely related to birds and, as a heterodontosaurid ornithischian, is on an entirely separate branch of the dinosaur family tree (Fig. 1).

And indeed, *Tianyulong* doesn't have true pennaceous feathers. It has long filaments, very similar to what have been called 'protofeathers'

or, more non-committally, 'dinofuzz'. These filaments are evident in some theropods such as *Caudipteryx* that have true pennaceous feathers, but are also found in a range of other theropods that lack definitive feathers, such as the basal coelurosaur *Sinosauropteryx*, the therizinosauroid *Beipiaosaurus* and the basal tyrannosauroid *Dilong*<sup>3,4,10</sup>.

Herein lies the controversy. No one disputes that these filaments are integumentary (in the skin). The question is, from what part of the

skin do they come — the outside or the inside? Many<sup>3,4,10,11</sup> have regarded them as epidermal (that is, as projecting, external appendages that are somehow evolutionarily related to feathers). Others<sup>5,6</sup> have regarded them as dermal (that is, as the remains of collagen fibres below the skin's surface). In this context, the difference between epidermal and dermal is huge. If they are epidermal, then they bear not only on feather evolution and avian origins, but also on metabolic physiology, behavioural display and flight. If they are dermal, then they're ultimately structural in function and have little bearing on those other issues. Unfortunately, the rhetoric of this debate has overshadowed the scientific evaluation of evidence<sup>9</sup>. If the integumentary filaments of *Tianyulong* are dermal (collagen fibres), then they become interesting but not of monumental importance. However, if they are epidermal, then they take on great significance.

Given the position of *Tianyulong* near the evolutionary base of ornithischian dinosaurs, the presence of epidermal, filamentous, feather-like structures could mean that the ancestral dinosaur was a fuzzy (though maybe not cuddly) animal. Of course, that would also mean that a fuzzy coat of protofeathers was lost many times in dinosaur evolution, because lots of dinosaur groups on both great branches of the dinosaur family tree are known to have scaly, reptilian skin (Fig. 1). But, before complicated scenarios for feather evolution are concocted, the fundamental question to be answered is whether the filaments of *Tianyulong* are on the outside or inside of the skin's surface.

That seemingly simple question is surprisingly hard to answer. The obvious test would be a biochemical or molecular assay to find out if the filaments are composed of the feather protein keratin or the collagen protein, but the mode of fossil preservation may not

## CHEMISTRY

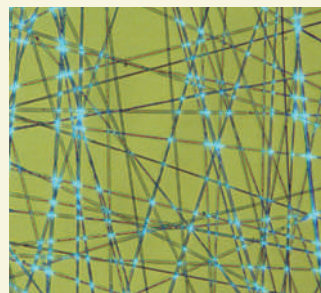
# Thinking outside the flask

The present credit crunch is forcing everyone to save money, and chemists are no exception. A good cost-cutting measure is to perform reactions on a small scale, thereby reducing the outlay on raw materials and minimizing the energy required to drive the reactions. Reporting in *Nature Chemistry*, Anzenbacher and Palacios describe the ultimate in miniature reaction vessels — junctions formed when two polymer nanofibres are fused together (P. Anzenbacher Jr & M. A. Palacios *Nature Chem.* doi:10.1038/nchem.125; 2009).

The authors prepared nanofibres — each hundreds of times narrower than a human hair — from readily

available polymers, and loaded them with various chemical reactants. They then laid fibres containing different reactants across each other and exposed them to either heat or solvent vapour. This caused the fibres to fuse together, forming junctions that defined discrete chemical reactors with attolitre-scale volumes (1 attolitre is  $10^{-18}$  litres). The junctions contained as few as 1,500 molecules of each reactant.

To illustrate the principle of their ultra-small reactors, Anzenbacher and Palacios doped fibres with two non-fluorescent compounds that form a fluorescent product when they react. When the two types of fibres were overlapped to form



a random mat and then heated, fluorescence at the fused junctions clearly indicated the formation of the reaction product (pictured).

The authors showed that several types of reactions can be performed in their attoreactors, including those in which one of the reactants is polymeric (although reactions with two polymeric reactants are expected to be problematic,

because polymers can't easily diffuse through the nanofibre matrix). Furthermore, the products can be analysed directly within the junctions using fluorescence measurements or mass spectrometry.

Several applications for these attoreactors suggest themselves. For example, libraries of nanofibres could be prepared in which each nanofibre is loaded with a different compound from the same chemical class. Selected libraries could then be reacted with each other, as in combinatorial chemistry, to prepare many different products quickly and easily. If the products could be screened directly for biological activity, this would be useful for the high-throughput preparation and testing of compounds in drug-discovery programmes.

**Stephen Davey**



permit a meaningful test. That leaves observable structure, and here the best indicator would be whether the filaments are tubular or solid in cross-section. Feathers are hollow, tubular structures, and developmental models predict that 'protofeather' filaments should also be hollow<sup>11</sup>. By contrast, collagenous filaments should be essentially solid.

Zheng and colleagues<sup>7</sup> interpret the filaments of *Tianyulong* as being hollow, although it's fair to question their evidence (longitudinal stripes on the filaments). But other attributes of the filaments are also highly suggestive of their being epidermal. For example, the filaments associated with the base of the tail are extremely long, and, given that the tail is already reinforced internally with stiffening rods of ossified tendons, it is possible that these filaments indeed project outside the skin's surface. Even if they can be shown to be definitively epidermal, the ultimate question is whether they are

part of the evolutionary lineage of true feathers or an independent evolution of projecting epidermal appendages. Certainly, the finding of very differently structured, projecting, hollow 'bristles' on the tail of another Yixian ornithischian, *Psittacosaurus* (a basal horned dinosaur)<sup>12</sup>, raises the possibility that there may be a range of filamentous epidermal structures in dinosaurs, and that not all such structures may be related evolutionarily to feathers<sup>13</sup>.

Perhaps the only clear conclusion that can be drawn from the foregoing is that little *Tianyulong* has made an already confusing picture of feather origins even fuzzier. Such an outcome is common in palaeontology. But the prospects of new fossils, new molecular and imaging techniques (such as synchrotron tomography), and even new ideas, offer the hope of bringing the evolutionary picture into sharper focus — and that picture may well end up being of fuzzy dinosaurs. ■

Lawrence M. Witmer is in the Department of Biomedical Sciences, Ohio University College of Osteopathic Medicine, Athens, Ohio 45701, USA. e-mail: witmerl@ohio.edu

1. Zhou, Z., Barrett, P. M. & Hilton J. *Nature* **421**, 807–814 (2003).
2. Norell, M. A. *et al.* *Nature* **416**, 36–37 (2002).
3. Norell, M. A. & Xu, X. *Annu. Rev. Earth Planet. Sci.* **33**, 277–299 (2005).
4. Zhang, F., Zhou, Z. & Dyke, G. *Geol. J.* **41**, 395–404 (2006).
5. Feduccia, A. *et al.* *J. Morphol.* **266**, 125–166 (2005).
6. Lingham-Soliar, T., Feduccia, A. & Wang, X. *Proc. R. Soc. Lond. B* **274**, 1823–1829 (2007).
7. Zheng, X.-T., You, H.-L., Xu, X. & Dong, Z.-M. *Nature* **458**, 333–336 (2009).
8. Butler, R. J., Upchurch, P. & Norman, D. B. J. *Syst. Palaeontol.* **6**, 1–40 (2008).
9. Witmer, L. M. in *Mesozoic Birds* (eds Chiappe, L. M. & Witmer, L. M.) 3–30 (Univ. California Press, 2002).
10. Chen, P.-J., Dong, Z.-m. & Zhen, S.-n. *Nature* **391**, 147–152 (1998).
11. Prum, R. O. & Brush, A. H. Q. *Rev. Biol.* **77**, 261–295 (2002).
12. Mayr, G. *et al.* *Naturwissenschaften* **89**, 361–365 (2002).
13. Sawyer, R. H., Washington, L. D., Salvatore, B. A., Glenn, T. C. & Knapp, L. W. *J. Exp. Zool.* **297B**, 27–34 (2003).

## GLOBAL CHANGE

# West-side story of Antarctic ice

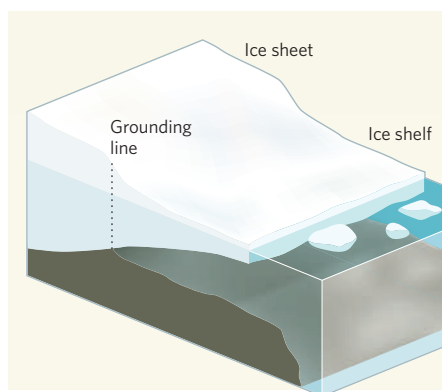
Philippe Huybrechts

**During the past five million years, the West Antarctic ice sheet has waxed and waned in size. A two-pronged reconstruction of that history provides clues to the ice sheet's future behaviour.**

For human societies, the prospect of sea-level rise is probably the most serious long-term threat from unabated climate warming. Both the Greenland ice sheet and the West Antarctic ice sheet are believed to be vulnerable to warming at the levels projected for the coming centuries. Sustaining those levels for many more centuries or millennia could ultimately cause the demise of both ice sheets — producing a worldwide rise in sea level of about 12 metres compared with today's levels, of which some 5 metres would derive from West Antarctica<sup>1</sup>.

The past can be a guide to the future. But the late Quaternary (the past 0.5 million to 1 million years), for which we have the most data, was generally colder than today, and so is not an ideal analogue for the future. The twin papers by Naish *et al.*<sup>2</sup> and Pollard and DeConto<sup>3</sup>, published in this issue, now present a step towards filling this gap in data and understanding. In their reconstruction of Antarctica's glacial history, the authors concentrate on the early to middle Pliocene, an interval of time between 5 million and 3 million years ago, when planetary temperatures were more in the range of those projected for the coming centuries.

The suspected vulnerability of the West Antarctic ice sheet stems from its particular setting. It is grounded mostly below sea level and is surrounded by large floating ice shelves (Fig. 1). These floating extensions are in direct contact



**Figure 1 | The West Antarctic ice sheet.** This much simplified depiction shows, to the left, the grounded marine ice, which sits upon bedrock or sediment on the sea floor. To the right, separated by the 'grounding line', is the floating ice shelf, which is thought to buttress the grounded ice sheet. The new observational study<sup>2</sup> and modelling work<sup>3</sup> identify a 40,000-year cycle in which the grounding line moves back and forth across the sea floor between glacial and interglacial states, punctuated by periodic ice-sheet collapses with more or less open-ocean conditions during super-interglacials.

with the ocean and are widely believed to have a crucial role in keeping the grounded ice sheet in place. At present, Antarctic temperatures are too low to generate significant surface

melting during the summer, and this will remain so even with moderate warming. Some combination of sea-level changes and changes in sub-ice-shelf melting must therefore be a key process by which variations in global climate control changes in the West Antarctic ice sheet. A long-standing debate among glaciologists has been the mechanism by which changes to the ice shelf are transmitted inland to grounded ice sheets across their common boundary, the 'grounding line'<sup>4</sup> (Fig. 1), and how effective this mechanism might be.

Naish and colleagues' observational evidence (page 322)<sup>2</sup> comes from a new sediment core recovered from beneath the Ross Ice Shelf near the current ice-shelf margin. The site turned out to be a strategic location. The range of sediments and rock types and other features seen in the core allowed the authors to distinguish between a cyclical succession of conditions at the drill location: open ocean with little or no summer sea-ice (super-interglacial conditions warmer than those of today); coverage by a floating ice shelf (interglacial conditions much like today's); and overriding by an ice sheet grounded on the sea floor (glacial conditions). Moreover, unlike many other glacial records in which each glaciation has erased the evidence of the previous one, this record stands out in having large sections without such hiatuses because tectonic forces have produced a favourable rate of bedrock sinking.

Dating of undisturbed sections along the core enabled Naish *et al.* to identify 40 sedimentary cycles, each of about 40,000 years' duration, during the Pliocene (up to 1.8 million years ago). These results are in good accordance with the same cyclicity seen in marine-isotope records of global ice volume and mean deep-sea temperatures<sup>5</sup>, and are in phase with summer half-year insolation caused by variations of the Earth's tilt at the same periodicity.

The geological sediment data, however, can only provide information on variations of the extent of the grounded ice sheet at one particular location. They provide no indication of ice thickness or, by extension, ice volume, which is where the modelling comes in.

Pollard and DeConto (page 329)<sup>3</sup> used a three-dimensional ice-sheet model to simulate the evolution of the West Antarctic ice sheet in a manner consistent with the geological data. A novel feature of their ice-flow model is the way the different flow regimes of grounded and floating ice are coupled across the grounding line, following the results of recent theoretical work<sup>6</sup>. Although approximations remain, the incorporation of this work in a three-dimensional model is new and is a notable methodological contribution to one of glaciology's grand unsolved problems — the mechanics of how the West Antarctic grounding line migrates over time<sup>4</sup>.

The model confirms that the conditions reflected in the sediment core are probably indicative of the evolution of the West Antarctic ice sheet as a whole during the past 5 million years. In addition, the model roughly reproduces much of the inferred ice-sheet variations, ranging from full glacial extents, to intermediate interglacial states similar to those pertaining today, to brief collapses of the entire West Antarctic ice sheet during the warmest super-interglacials. Unlike Antarctic ice-sheet reconstructions produced with broadly similar models, including my own<sup>7,8</sup>, Pollard and DeConto find a much more prominent role for sub-ice-shelf melting than for global sea-level variation in driving past migrations of West Antarctic grounding lines.

However, the modelling does not solve every problem. The episodic collapses of the West Antarctic ice sheet during the past 5 million years also happen at times when there is little or no evidence for open-ocean conditions or high sea levels, adding an erratic component to the simulations. Other aspects — such as the timing of grounding-line retreat in the Ross Sea sector since the Last Glacial Maximum, and probably unduly thick ice in this sector both then<sup>9</sup> and today — seem to contradict the observational evidence. This calls for further refinements of processes, such as the sliding of the ice over its bed, that will increase the credibility of the model for supporting predictions on shorter timescales. Other issues concern the rather simple schemes used to prescribe crucial model forcing, such as sub-ice-shelf melting. These would best be addressed by more detailed global-climate and ocean models than are currently possible on these long timescales.

Even so, the modelling by Pollard and DeConto<sup>3</sup> may already be robust enough to attempt to put initial constraints on the amount of nearby ocean warming required to generate enough sub-ice-shelf melting to initiate a significant retreat of the West Antarctic ice sheet. The required ocean warmings, of the order of 5 °C, may well take several centuries to develop.

But such an outcome could result from the accumulation of total greenhouse-gas emissions projected for the twenty-first century, if emissions are not greatly reduced<sup>1</sup>. The implied transition time for a total collapse of the West Antarctic ice sheet of one thousand to several thousand years<sup>3</sup> seems rapid by Antarctic standards. But it is nowhere near the century timescales for West Antarctic ice-sheet decay based on simple marine ice-sheet models<sup>10</sup>. ■

Philippe Huybrechts is in the Earth System Sciences Group and the Department of Geography, Vrije Universiteit Brussel,

Pleinlaan 2, B-1050 Brussels, Belgium.

e-mail: phuybrec@vub.ac.be

1. Solomon, S. D. *et al.* (eds) *Climate Change 2007: The Physical Science Basis* (Cambridge Univ. Press, 2007).
2. Naish, T. *et al.* *Nature* **458**, 322–328 (2009).
3. Pollard, D. & DeConto, R. M. *Nature* **458**, 329–332 (2009).
4. Weertman, J. *Nature* **260**, 284–286 (1976).
5. Lisiecki, L. E. & Raymo, M. E. *Paleoceanography* doi:10.1029/2004PA001071 (2005).
6. Schoof, C. J. *Geophys. Res.* doi:10.1029/2006JF000664 (2007).
7. Ritz, C., Rommelaere, V. & Dumas, C. J. *Geophys. Res.* **106**, 31943–31964 (2001).
8. Huybrechts, P. *Quat. Sci. Rev.* **21**, 203–231 (2002).
9. Waddington, E. D. *et al.* *Geology* **33**, 281–284 (2005).
10. Thomas, R. H. J. *Glaciol.* **24**, 167–177 (1979).

## NEUROSCIENCE

# Secret of synapse specificity

Scott M. Thompson and Hayley A. Mattison

**How does the brain organize all of the information stored in memory? On the basis of a state-of-the-art imaging study of neuronal activity in real time, the answer seems to be, through specificity in space and time.**

Memories are encoded by a specific pattern of activity that is unique to the information being processed and stored. Memory formation is almost certainly achieved at the synaptic junctions between neurons through the process of long-term potentiation (LTP), whereby synaptic communication between two simultaneously active neurons becomes stronger<sup>1,2</sup>. One of the many attractions of LTP for explaining the biological basis of memory formation is that it displays remarkable synapse specificity<sup>3,4</sup>: only synapses that are activated become strengthened, and neighbouring synapses located only a micrometre or two away on the same neuron remain unaffected<sup>5</sup>. In an exciting paper on page 299 of this issue, Lee *et al.*<sup>6</sup> study living neurons using advanced fluorescence microscopy to conclusively demonstrate the molecular basis of the synapse specificity of LTP.

A crucial enzyme for triggering LTP is CaMKII. Release of glutamate, a neurotransmitter, from the presynaptic neuron leads to depolarization of the postsynaptic neuron, thus relieving the voltage-dependent block of the NMDA-type glutamate receptors on the postsynaptic neuron, and allowing calcium ions (Ca<sup>2+</sup>) to enter it. The increased intracellular Ca<sup>2+</sup> concentration leads to activation of CaMKII, which then phosphorylates itself, thereby remaining active even after the Ca<sup>2+</sup> concentration has fallen back to normal. Induction of LTP is therefore prevented either by NMDA-receptor antagonists or by CaMKII inhibitors<sup>7</sup>.

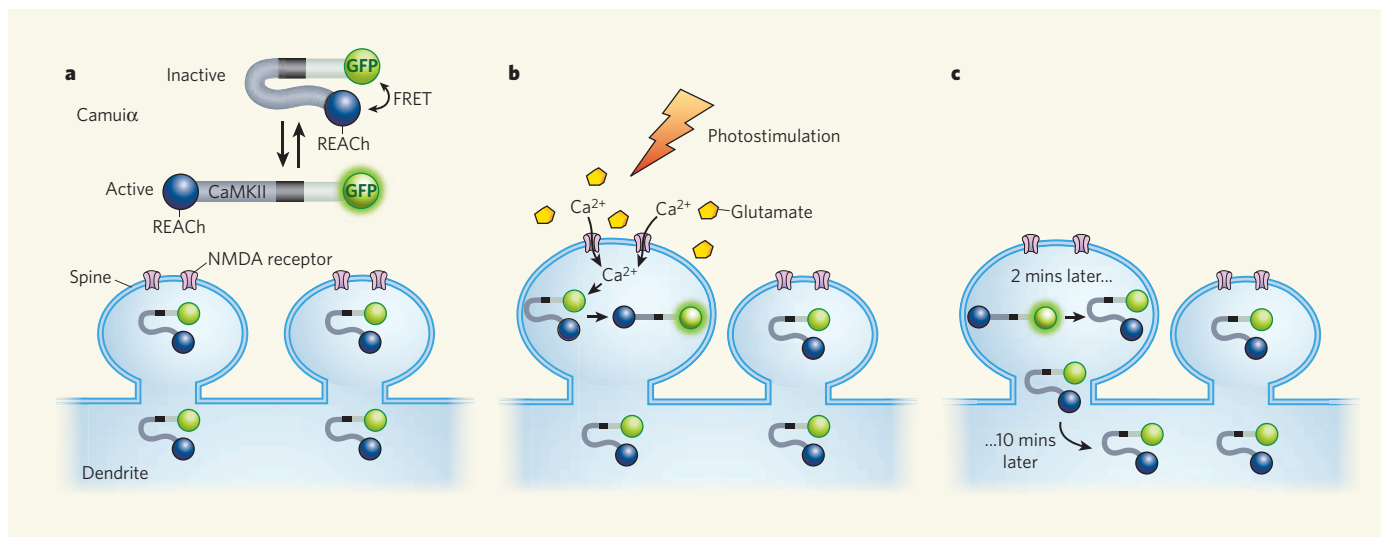
To image real-time activation of CaMKII in living neurons, Lee *et al.*<sup>6</sup> used the technique of fluorescence resonance energy transfer (FRET)<sup>8</sup>. In FRET, distance-dependent energy transfer from a donor fluorescent molecule

(fluorophore) to an acceptor fluorophore nanometres away is measured through changes in fluorescence intensity. Lee *et al.* modified a 'reporter' molecule — the CaMKII-based protein Camui<sup>9</sup> — that has a donor fluorophore attached at one end and an acceptor fluorophore at the other. On activation of CaMKII, the donor-acceptor pair moves apart, leading to decreased FRET. The authors further improved Camui<sup>9</sup> by using a bright fluorophore, green fluorescent protein (GFP), as the donor and a non-fluorescent acceptor fluorophore<sup>10</sup>, a mutated GFP called REACH. Thus, only changes in the emission of GFP need be monitored, allowing emitted photons to be collected in a wider spectral window.

Instead of measuring fluorescence intensity, Lee *et al.* measured the fluorescence lifetime of the donor fluorophore<sup>7</sup>. When CaMKII is inactive, REACH at one end of Camui<sup>9</sup> is close enough to quench GFP emission and shorten its fluorescence lifetime. With REACH moving away upon CaMKII activation, however, the fluorescence lifetime of GFP is prolonged (Fig. 1a). That may sound a simple concept, but the lifetime of GFP in the excited state is less than 2 nanoseconds. Extremely brief (<100-femtosecond) bursts of high-energy excitation light must therefore be delivered and high-efficiency detectors of photon emission must be used to measure the time between an excitation pulse and the arrival of an emitted photon.

Finally, to achieve adequate spatial resolution, as well as to detect as many emitted photons as possible, the authors used the technique of two-photon microscopy. This approach was particularly advantageous because of its superior spatial resolution, as the experiments had to be performed in relatively thick slices of





**Figure 1 | Synapse-specific long-term potentiation (LTP).** Lee *et al.*<sup>6</sup> used changes in the fluorescence lifetime of GFP bound to CamuIIa–CaMKII to study synapse-specific CaMKII activity. **a**, When CaMKII is inactive, the fluorescence lifetime of GFP is short. **b**, On photostimulation of NMDA receptors and subsequent  $\text{Ca}^{2+}$  influx, there is an increase both in GFP

fluorescence — corresponding to CaMKII activation — and in the volume of the stimulated spine through LTP. **c**, Although it takes 10 minutes for CaMKII to diffuse out of the spine, the enzyme becomes inactive within 2 minutes, so that it acts only within the stimulated dendritic spine, thus accounting for the synapse specificity of LTP.

brain tissue for LTP to be induced. The authors' idea was to induce LTP at a single dendritic spine — small protrusions on the dendritic processes of neurons that are sites of synaptic communication and that enlarge in response to LTP. To achieve such specificity, Lee *et al.* used photostimulation, whereby 'caged' glutamate molecules are locally released after exposure to high-energy light.

Having designed their tools so elegantly, Lee *et al.*<sup>6</sup> were ready to view CaMKII activation in single spines. Using patterns of photostimulation that induce LTP, they observed a sharp spike in the fluorescence lifetime of CamuIIa, corresponding to CaMKII activation. Importantly, the increase in fluorescence lifetime was confined only to the photostimulated spine, indicating that CaMKII activation is synapse specific (Fig. 1b). The authors could prevent the increase in CamuIIa's fluorescence lifetime using the NMDA-receptor antagonist AP5, thus showing that CaMKII activation was triggered by  $\text{Ca}^{2+}$  entry via NMDA receptors. Remarkably, although the structural and physiological changes underlying LTP persisted for hours, the increased lifetime of CamuIIa fluorescence — and so the period for which CaMKII remains active — lasted less than 2 minutes. This period of time is several orders of magnitude shorter than suggested from previous biochemical measurements, which had indicated that CaMKII remains active for hours after NMDA-receptor-mediated  $\text{Ca}^{2+}$  influx<sup>7</sup>.

Both  $\text{Ca}^{2+}$  and CaMKII can diffuse freely in the cytoplasm. So what accounts for the synapse specificity of this enzyme's activation? Why doesn't locally activated CaMKII diffuse to neighbouring synapses and trigger LTP there? With a single photostimulus, Lee *et al.*<sup>6</sup> observed that  $\text{Ca}^{2+}$  levels decay back to the resting concentration within a few hundred

milliseconds. Given the location of excitatory synapses on dendritic spines, this is too short a time for  $\text{Ca}^{2+}$  to diffuse to neighbouring spines and activate CaMKII there (Fig. 1c).

The authors found that the spread of activated CaMKII is limited not only by its rate of inactivation but also by its low mobility. Using a photoactivatable GFP variant as a fluorophore, they measured the diffusion of CaMKII out of the spine through fluorescence decay. After photoactivation, fluorescence declined as the enzyme diffused out of the synapse. The decay took 10 minutes, however, probably because CaMKII is anchored to other proteins in the spine<sup>7</sup>. The authors thus conclude that a fast rate of CaMKII inactivation, together with its much slower rate of diffusion, restricts CaMKII activity to the stimulated synapse (Fig. 1c).

Postsynaptic depolarization in the absence of glutamatergic stimulation can also activate CaMKII in the spine by opening L-type  $\text{Ca}^{2+}$  channels<sup>6</sup>. These channels, however, do not contribute to the depolarization-induced increase in bulk  $\text{Ca}^{2+}$  levels. The authors suggest provocatively that there may be distinct pools of CaMKII in the spine that are activated by  $\text{Ca}^{2+}$  influx through different routes. Only the CaMKII pool activated following NMDA-receptor-mediated  $\text{Ca}^{2+}$  influx can trigger LTP, perhaps because only those CaMKII molecules are anchored adjacent to appropriate substrates.

It is important to bear in mind that there may be conditions in which LTP does involve communication with adjacent synapses. The protocol Lee and colleagues used is somewhat different from the patterns of high-frequency stimulation (100 Hz) that are commonly used to induce LTP. Surprisingly, we know little about the spatial pattern of intracellular  $\text{Ca}^{2+}$  diffusion after such stimulation, let alone about

CaMKII activation. Intracellular  $\text{Ca}^{2+}$ -buffering molecules at the activated synapse are likely to become saturated by high-frequency stimulation patterns, leading to more  $\text{Ca}^{2+}$  escaping from the dendritic spine and, perhaps, to activation of CaMKII in neighbouring synapses. Indeed, intense activation of NMDA receptors leads to the spread of CaMKII activation along the stimulated dendrites of neurons maintained in culture<sup>11</sup>. Furthermore, induction of potentiation at one synapse is associated with an increased probability that synapses up to ten micrometres away from the stimulated synapse will themselves become potentiated, perhaps because of spread of the small GTPase enzyme Ras (ref. 5). It will be of interest to determine the synapse specificity of the key CaMKII substrates and Ras effectors that are ultimately responsible for long-term memory storage.

Scott M. Thompson and Hayley A. Mattison are in the Department of Physiology, University of Maryland School of Medicine, Baltimore, Maryland 21201, USA.  
e-mail: sthom003@umaryland.edu

1. Malenka, R. C. & Bear, M. F. *Neuron* **44**, 5–21 (2004).
2. Matylnia, A., Kushner, S. A. & Silva, A. J. *Annu. Rev. Genet.* **36**, 687–720 (2002).
3. Andersen, P., Sundberg, S. H., Svein, O. & Wigström, H. *Nature* **266**, 736–737 (1977).
4. Lynch, G. S., Dunwiddie, T. & Gribkoff, V. *Nature* **266**, 737–739 (1977).
5. Harvey, C. D., Yasuda, R., Zhong, H. & Svoboda, K. *Science* **321**, 136–140 (2008).
6. Lee, S.-J. R., Escobedo-Lozoya, Y., Szatmari, E. M. & Yasuda, R. *Nature* **458**, 299–304 (2009).
7. Lisman, J., Schulman, H. & Cline, H. *Nature Rev. Neurosci.* **3**, 175–190 (2002).
8. Yasuda, R. *Curr. Opin. Neurobiol.* **16**, 551–561 (2006).
9. Takao, K. *et al. J. Neurosci.* **25**, 3107–3112 (2005).
10. Ganesan, S. *et al. Proc. Natl Acad. Sci. USA* **103**, 4089–4094 (2006).
11. Rose, J., Jin, S. X. & Craig, A. M. *Neuron* **61**, 351–358 (2009).

# Hippopotamus and whale phylogeny

Arising from: J. G. M. Thewissen, L. N. Cooper, M. T. Clementz, S. Bajpai & B. N. Tiwari *Nature* **450**, 1190–1194 (2007)

Thewissen *et al.*<sup>1</sup> describe new fossils from India that apparently support a phylogeny that places Cetacea (that is, whales, dolphins, porpoises) as the sister group to the extinct family Raoellidae, and Hippopotamidae as more closely related to pigs and peccaries (that is, Suina) than to cetaceans. However, our reanalysis of a modified version of the data set they used<sup>2</sup> differs in retaining molecular characters and demonstrates that Hippopotamidae is the closest extant family to Cetacea and that raoellids are the closest extinct group, consistent with previous phylogenetic studies<sup>2,3</sup>. This topology supports the view that the aquatic adaptations in hippopotamids and cetaceans are inherited from their common ancestor<sup>4</sup>.

To conduct our analyses, we started with the same published matrix that Thewissen *et al.* modified<sup>2</sup>, but included all taxa (but see Appendix Fig. 2 where *Andrewsarchus* is removed from our results), retained the molecular partition that was derived from an earlier study<sup>5</sup>, added the three anthracotheres included by Thewissen *et al.*<sup>1</sup>, replaced 'Raoellidae' with their character codings for *Indohyus* and *Khirtharia*, included most of the changes and additions to pakicetids that Thewissen *et al.*<sup>1</sup> suggested (see Appendix), and then conducted cladistic analyses of the entire matrix, with all characters weighted equally<sup>6</sup> (see Appendix), as well as with homoplastic characters down-weighted<sup>7</sup>. We consider it particularly important to include molecular data because they are widely recognized as being instrumental in reconstructing mammalian phylogeny<sup>8</sup>.

In the equal weights and differential weights analyses, Hippopotamidae is the closest extant taxon to Cetacea, and in those that include

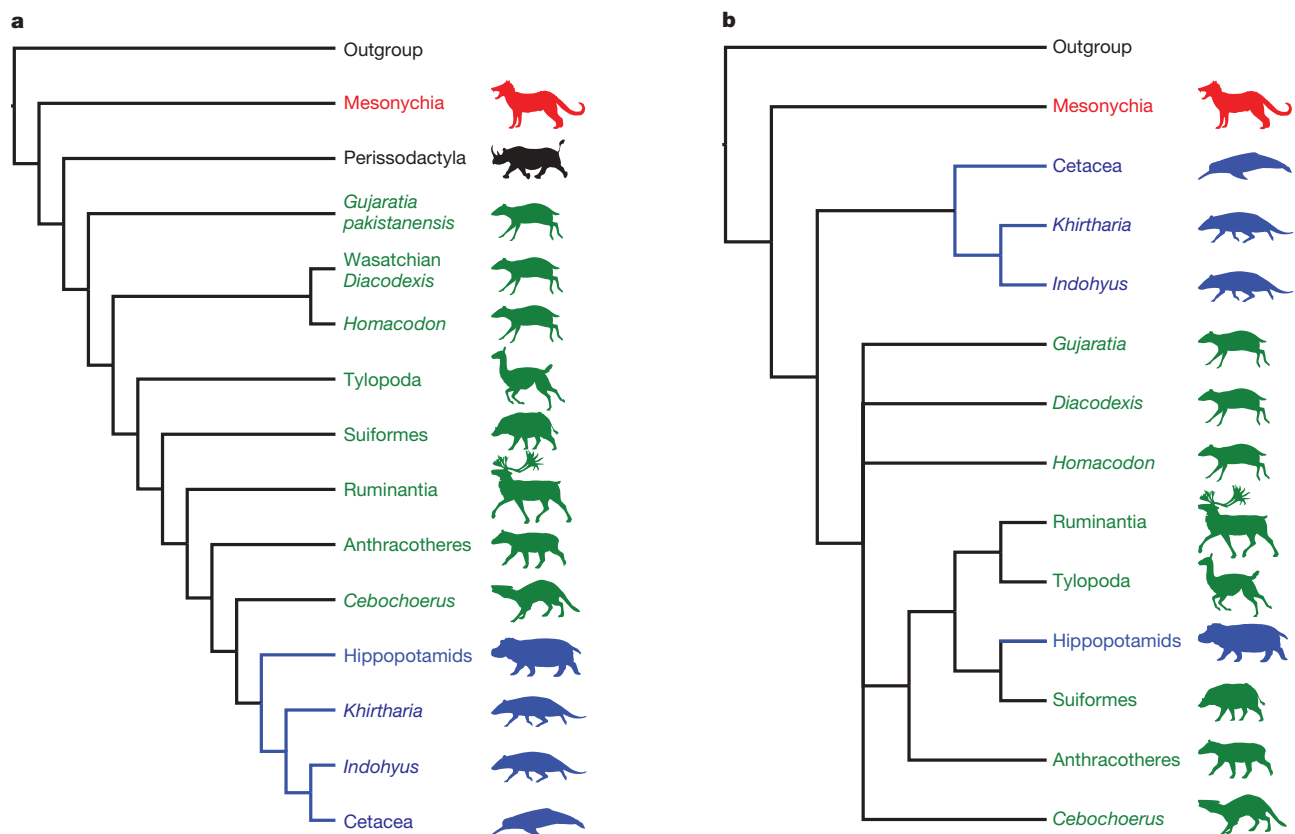
fossils, Raoellidae or the raoellid *Indohyus* is more closely related to Cetacea than is Hippopotamidae (Fig. 1). Hippopotamidae is the exclusive sister group to Cetacea plus Raoellidae in the analysis that down-weights homoplastic characters, although in the equally weighted analysis, another topology was equally parsimonious. In that topology, Hippopotamidae moved one node out, being the sister group to an *Andrewsarchus*, Raoellidae and Cetacea clade. In neither analysis is Hippopotamidae closer to the pigs and peccaries than to Cetacea, the result obtained by Thewissen *et al.*<sup>1</sup>. In all our analyses, pachyostosis (thickening) of limb bones and bottom walking, which occur in hippopotamids<sup>9,10</sup>, are interpreted to have evolved before the pachyostosis of the auditory bulla, as seen in raoellids and cetaceans<sup>1</sup>. Pachyostosis of the bulla, referred to as the involucrum, is thought to be an adaptation for underwater hearing<sup>11</sup>. Thus, by looking at the limited pachyostosis in hippopotamids, we can speculate that pachyostosis in the ancestors of whales initially evolved to counteract buoyancy, but was later co-opted for underwater hearing as it spread to the cranium. This and other hypotheses like it are tenable using our trees, but not those of Thewissen *et al.*<sup>1</sup>, and underscore the importance of including different classes of data in phylogenetic analyses.

Jonathan H. Geisler<sup>1</sup> & Jessica M. Theodor<sup>2</sup>

<sup>1</sup>Department of Geology and Geography and Georgia Southern Museum, Georgia Southern University, Statesboro, Georgia 30460-8149, USA.

e-mail: geislerj@georgiasouthern.edu

<sup>2</sup>Department of Biological Sciences, University of Calgary, Calgary, Alberta T2N 1N4, Canada.



**Figure 1 | Effect of molecular data on the phylogeny of cetaceans and terrestrial relatives.** **a**, Simplified version of the single most parsimonious tree from the present study when homoplastic characters are down-weighted. **b**, Simplified version of the strict consensus shown in Fig. 2 of Thewissen *et al.*<sup>1</sup>. The carnivorous to omnivorous mesonychians are in red, terrestrial

cetartiodactyls (even-hoofed ungulates) are in green, and aquatic to semi-aquatic cetartiodactyls are in blue. Branches are coloured (blue) to show when aquatic adaptations, such as bottom walking, evolved. In **a** these features may have evolved earlier, depending on the behaviour of anthracotheres and cebochoerids. Details of the tree structure in **a** are available in the Appendix.



Received 26 March; accepted 15 October 2008.

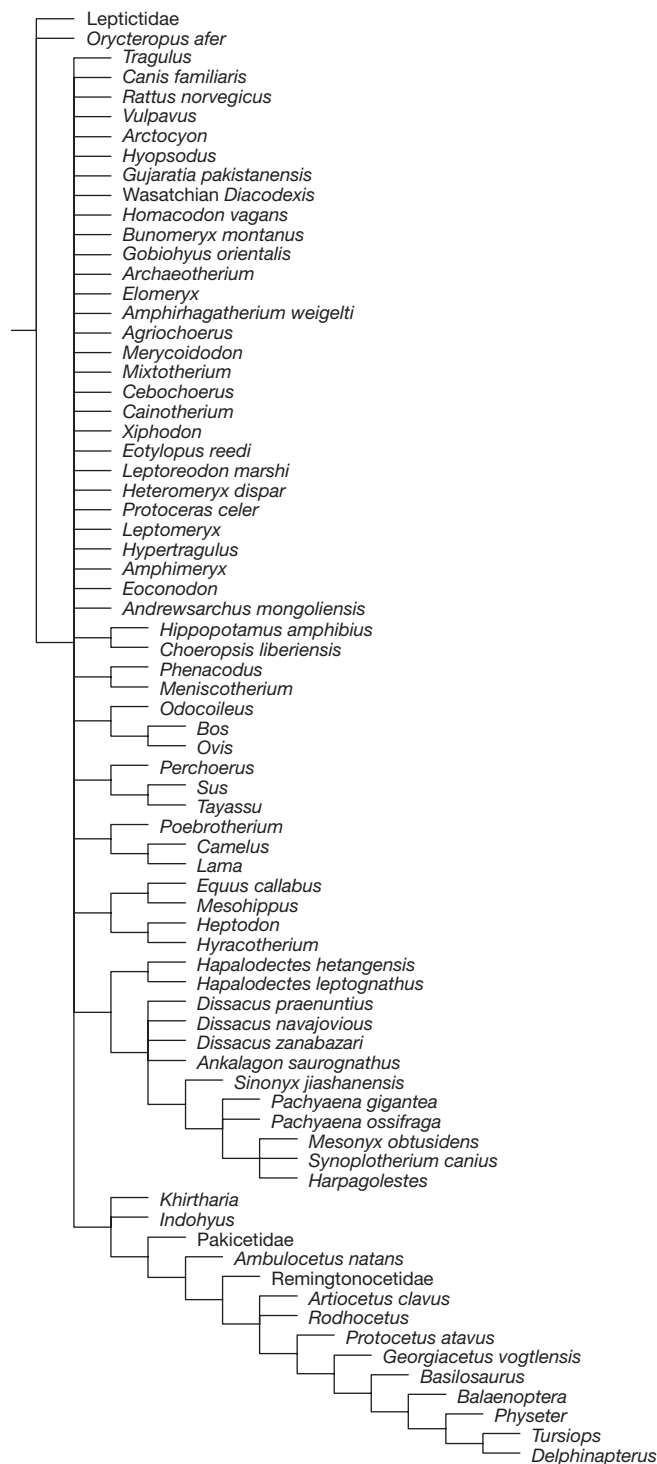
1. Thewissen, J. G. M., Cooper, L. N., Clementz, M. T., Bajpai, S. & Tiwari, B. N. Whales originated from aquatic artiodactyls in the Eocene epoch of India. *Nature* **450**, 1190–1194 (2007).
2. Geisler, J. H. & Uhen, M. D. Phylogenetic relationships of extinct Cetartiodactyls: results of simultaneous analyses of molecular, morphological, and stratigraphic data. *J. Mamm. Evol.* **12**, 145–160 (2005).
3. Geisler, J. H., Theodor, J. M., Uhen, M. D. & Foss, S. E. in *The Evolution of Artiodactyls* (eds Prothero, D. R. & Foss, S. E.) 19–31 (John Hopkins Univ. Press, 2007).
4. Gatesy, J., Hayashi, C., Cronin, M. A. & Arctander, P. Evidence from milk casein genes that cetaceans are close relatives of hippopotamid artiodactyls. *Mol. Biol. Evol.* **13**, 954–963 (1996).
5. Gatesy, J., Matthee, C., DeSalle, R. & Hayashi, C. Resolution of the supertree/supermatrix paradox. *Syst. Biol.* **51**, 652–664 (2002).
6. Goloboff, P., Farris, S. & Nixon, K. TNT (Tree Analysis Using New Technology) (Willi Hennig Society Edition, 2000).
7. Goloboff, P. A. Estimating character weights during tree search. *Cladistics* **9**, 83–91 (1993).
8. Springer, M. S., Stanhope, M. J., Madsen, O. & de Jong, W. W. Molecules consolidate the placental mammal tree. *Trends Ecol. Evol.* **19**, 430–438 (2004).
9. Wall, W. P. The correlation between high limb-bone density and aquatic habitats in recent mammals. *J. Paleontol.* **57**, 197–207 (1983).
10. Eltringham, S. K. *The Hippos* (Academic Press, 1999).
11. Nummela, S., Thewissen, J. G. M., Bajpai, S., Hussain, T. & Kumar, K. Sound transmission in archaic and modern whales: anatomical adaptations for underwater hearing. *Anat. Rec.* **290**, 716–733 (2007).

doi:10.1038/nature07776

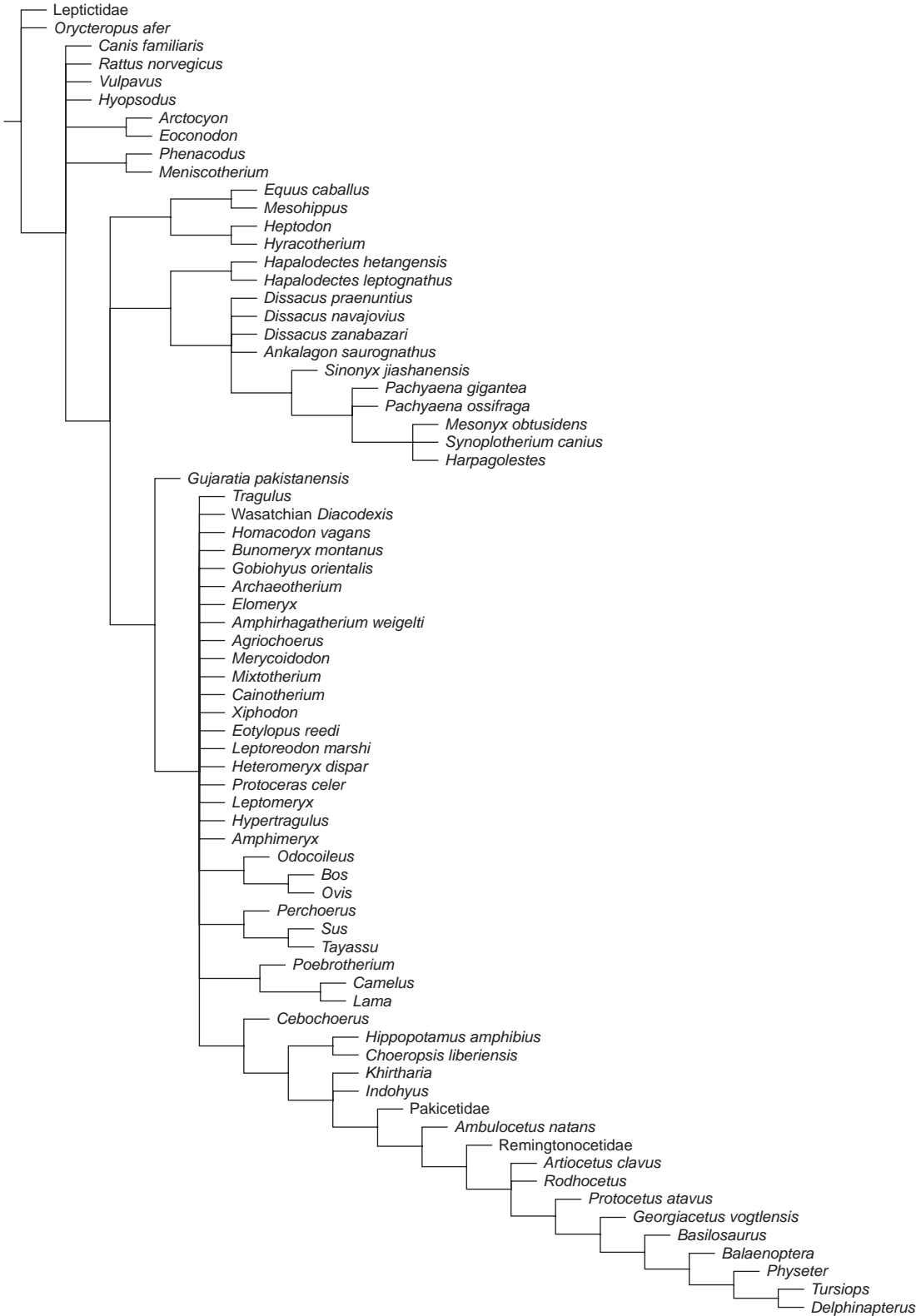
## APPENDIX

Three character codings for pakicetids from Thewissen *et al.*<sup>1</sup> were not followed here: character 38 was coded as state 1, character 142 also as state 1, and character 105 as state 2. These character codings are based on J.H.G.'s examination of pakicetid specimens currently housed at the Department of Anatomy, Northeastern Ohio Universities College of Medicine. The entire matrix used in the reanalysis and the tree files are available at <http://www.cetacea-evolution.org/main/>, which is supported by National Science Foundation grant DEB-0640361.

Phylogenetic analyses were conducted with the computer application TNT<sup>6</sup>. Two separate analyses were conducted: one with all characters given equal weight (Appendix Figs 1 and 2) and another that uses implied weighting (Appendix Fig. 3), which dynamically down-weights characters that show greater homoplasy during the analysis. The degree of down-weighting is set by the constant  $k$ , and the default value of 3 was used. Both analyses used the default settings for a 'New Technology Search', which uses sectorial searches and tree fusing. Exceptions to the defaults are that the heuristic searches were terminated when the minimum length was found 1,000 times and the maximum number of trees saved was set to 10,000.



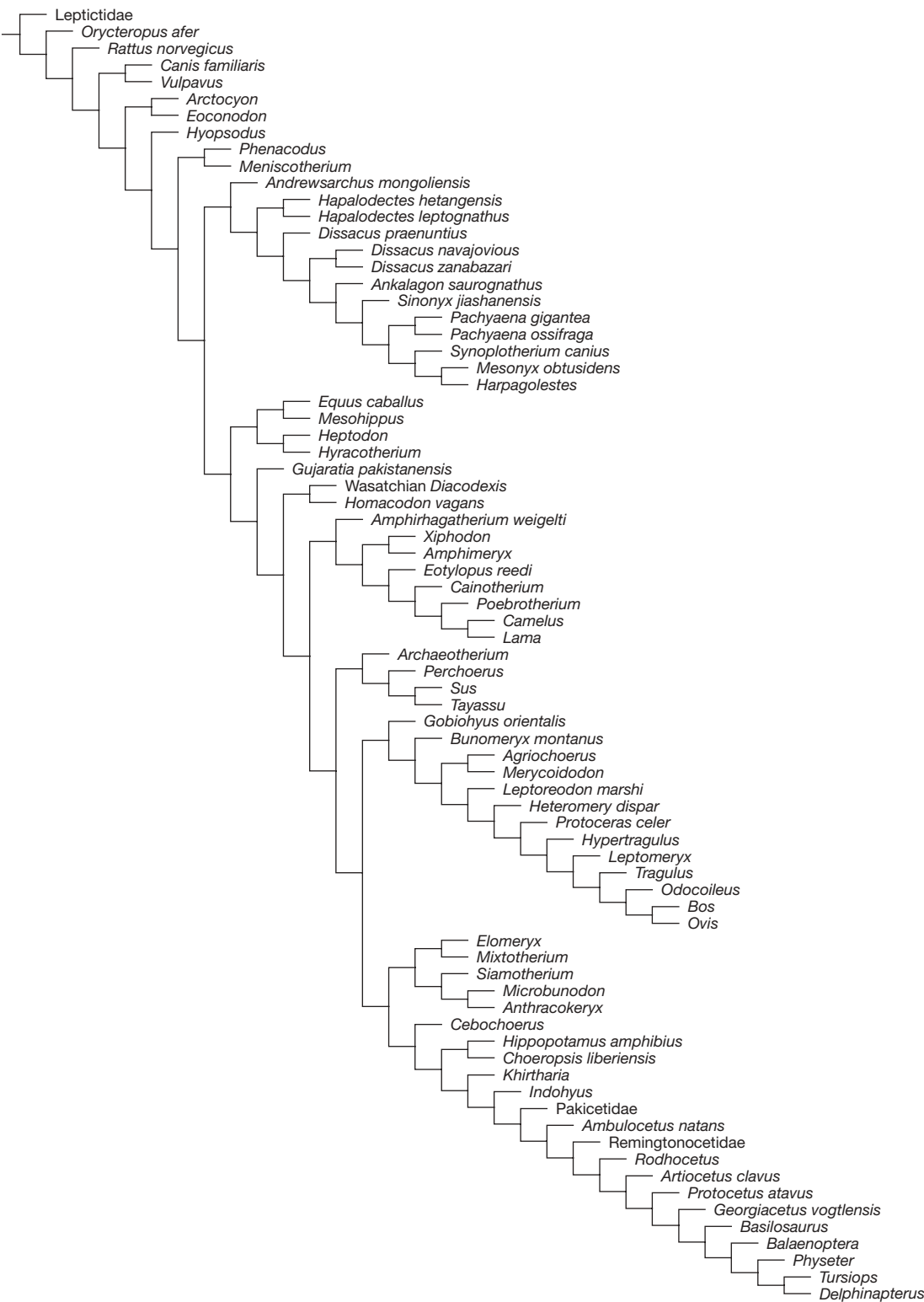
**Appendix Figure 1 | Strict consensus of equal-weights parsimony analysis.** A total of 1,584 most parsimonious trees of 33,490 steps was found. The tree files can be found at <http://www.cetacea-evolution.org/main/> and the strict consensus of these trees is shown.



**Appendix Figure 2 | Reduced consensus of equal-weights parsimony analysis.** Much of the lack of resolution in the strict consensus (Fig. 1) can be attributed to differing positions for the extinct taxon *Andrewsarchus*. If it

is pruned from all 1,584 most parsimonious trees, then the reduced consensus is as shown.





**Appendix Figure 3 | Differential weights parsimony analysis.** The analysis using implied weights with  $k = 3$  found only one optimal tree, which is shown. This tree is the basis for the simplified tree in the main text and has a fit of 2329.20679.

# Thewissen et al. reply

Replying to: J. H. Geisler & J. M. Theodor *Nature* 458, doi:10.1038/nature07776 (2009)

The analysis of Geisler and Theodor<sup>1</sup> confirms our main phylogenetic result<sup>2</sup>, that raoellids are, or include, the sister group to cetaceans. Their study expands on our findings by inferring that hippopotamids are the sister group to the combined raoellid–cetacean clade, whereas our paper had explicitly stated that our data could not address the position of the extant artiodactyl families.

Geisler and Theodor<sup>1</sup> place confidence in their results with regard to hippopotamids by stating that their analysis is consistent with “previous phylogenetic studies”. However, one of the two articles that they cite<sup>3</sup> was published after the publication of our paper, and they do not cite a recent paper<sup>4</sup> that disagrees with their (and our own) results.

Geisler and Theodor<sup>1</sup> execute two analyses, one with equally weighted characters and one with down-weighted characters (Appendix Fig. 3 of their paper). They state that in both analyses Hippopotamidae is the closest extant taxon to Cetacea, but their Appendix is inconsistent with this statement. In the Appendix, a large polychotomy lies at the base of the raoellid–cetacean clade in the equally weighted analysis. This polychotomy dissolves when *Andrewsarchus* is excluded from the study, which results in a strict consensus tree where hippopotamids are indeed the sister group to the raoellid–cetacean clade. Instead of discussing this consensus tree, Geisler and Theodor<sup>1</sup> mention that the equally weighted analysis yielded two most parsimonious trees: one showing hippopotamids as the sister group to the raoellid–cetacean clade and the other showing (apparently undeleted) *Andrewsarchus* as more closely related to the raoellid–cetacean clade than to hippopotamids. Notably, it took the deletion of *Andrewsarchus* before hippopotamids could be identified as the raoellid–cetacean sister group in the unweighted analysis of the Appendix of Geisler and Theodor<sup>1</sup>; the text seems to identify *Andrewsarchus* as a possible close relative of hippopotamids and this is consistent with another recently published cladogram<sup>4</sup>. *Andrewsarchus* is poorly known, and this may result in its unstable phylogenetic position, although it may still be part of the hippopotamid–raoellid–cetacean clade.

Geisler and Theodor<sup>1</sup> agree with us that the phylogenetic link between cetaceans and raoellids is closer than that between cetaceans and hippopotamids, and we consider the former link to be more robust than the latter. We also believe that improving fossil collections of poorly known taxa is important in advancing understanding of cetacean relationships. Hippopotamids are certainly closely related to cetaceans, yet they appear about 35 million years after the first cetaceans, and their origins are in Africa, not the birthplace of cetaceans.

Geisler and Theodor<sup>1</sup> further elaborate on our results by discussing the origin of osteosclerosis of limbs and bulla, calling this process pachyostosis (for definitions, see refs 5, 6). Osteosclerosis, but not pachyostosis, of the limbs occurs in hippopotamids, raoellids and pakicetids, whereas pachyosteosclerosis of the bulla occurs in raoellids and pakicetids, but not in hippopotamids. Experimental studies with lipoprotein-receptor-related proteins in transgenic mice suggest that pachyostosis and osteosclerosis are underlain by different genetic mechanisms<sup>7</sup>. The mechanisms that lead to such hyperostoses deserve further study, and such a study is necessary to test the speculation of Geisler and Theodor<sup>1</sup> that hyperostoses of the limbs and bulla are related. Such a study should be rooted in a sound understanding of the developmental controls of bone formation, a field that is growing rapidly<sup>8</sup> and that promises to elucidate cetacean evolution considerably.

**J. G. M. Thewissen<sup>1</sup>, Lisa Noelle Cooper<sup>1,2</sup>, Mark T. Clementz<sup>3</sup>, Sunil Bajpai<sup>4</sup> & B. N. Tiwari<sup>5</sup>**

<sup>1</sup>Department of Anatomy, Northeastern Ohio Universities College of Medicine, Rootstown, Ohio 44272, USA.

e-mail: thewisse@neoucom.edu

<sup>2</sup>School of Biomedical Sciences, Kent State University, Kent, Ohio 44242, USA.

<sup>3</sup>Department of Geology and Geophysics, University of Wyoming, Laramie, Wyoming 82071, USA.

<sup>4</sup>Department of Earth Sciences, Indian Institute of Technology, Roorkee 247 667, Uttarakhand, India.

<sup>5</sup>Wadia Institute of Himalayan Geology, G.M.S. Road, Dehra Dun 248 001, Uttarakhand 248 001, India.

1. Geisler, J. H. & Theodor, J. M. Hippopotamus and whale phylogeny. *Nature* 458, doi:10.1038/nature07776 (2009).
2. Thewissen, J. G. M., Cooper, L. N., Clementz, M. T., Bajpai, S. & Tiwari, B. N. Whales originated from aquatic artiodactyls in the Eocene epoch of India. *Nature* 450, 1190–1194 (2007).
3. Geisler, J. H., Theodor, J. M., Uhen, M. D. & Foss, S. E. in *The Evolution of Artiodactyls* (eds Prothero, D. R. & Foss, S. E.) 19–31 (Johns Hopkins Univ. Press, 2007).
4. O’Leary, M. A. & Gatesy, J. Impact of increased character sampling on the phylogeny of Cetartiodactyla (Mammalia): combined analysis including fossils. *Cladistics* 24, 397–442 (2008).
5. Francillon-Vieillot, H. et al. in *Skeletal Biomineralization Patterns, Processes, and Evolutionary Trends* (ed. Carter, J. G.) 471–530 (Van Nostrand Reinhold, 1990).
6. de Ricqlès, A. & de Buffrénil, V. in *Secondary Adaptations of Tetrapods to Life in Water* (eds Mazin, J. M. & Buffrénil, V. de) 289–310 (Verlag Dr. Friedrich Pfeil, 2001).
7. Bodine, P. V. N. & Komm, B. S. Wnt signaling and osteoblastogenesis. *Rev. Endocr. Metab. Disord.* 7, 33–39 (2006).
8. Hall, B. K. *Bones and Cartilage, Developmental and Evolutionary Skeletal Biology* 760 (Elsevier, 2005).

doi:10.1038/nature07775



# Activation of CaMKII in single dendritic spines during long-term potentiation

Seok-Jin R. Lee<sup>1</sup>, Yasmin Escobedo-Lozoya<sup>1</sup>, Erzsebet M. Szatmari<sup>1</sup> & Ryohei Yasuda<sup>1</sup>

**Calcium/calmodulin-dependent kinase II (CaMKII) plays a central part in long-term potentiation (LTP), which underlies some forms of learning and memory. Here we monitored the spatiotemporal dynamics of CaMKII activation in individual dendritic spines during LTP using two-photon fluorescence lifetime imaging microscopy, in combination with two-photon glutamate uncaging. Induction of LTP and associated spine enlargement in single spines triggered transient (~1 min) CaMKII activation restricted to the stimulated spines. CaMKII in spines was specifically activated by NMDA receptors and L-type voltage-sensitive calcium channels, presumably by nanodomain  $\text{Ca}^{2+}$  near the channels, in response to glutamate uncaging and depolarization, respectively. The high degree of compartmentalization and channel specificity of CaMKII signalling allow stimuli-specific spatiotemporal patterns of CaMKII signalling and may be important for synapse-specificity of synaptic plasticity.**

CaMKII is a serine/threonine protein kinase consisting of 12 subunits<sup>1</sup>. Each subunit is activated by the association of  $\text{Ca}^{2+}$ -bound calmodulin ( $\text{Ca}^{2+}/\text{CaM}$ ). When a CaMKII subunit is autophosphorylated at site T286, its activity can be maintained after the dissociation of  $\text{Ca}^{2+}/\text{CaM}$  (ref. 2). It has been suggested that CaMKII activation in postsynaptic density may persist long term (more than hours) to maintain LTP<sup>2</sup>. CaMKII interacts with several ion channels including the NR2B subunit of NMDA receptors (NMDARs)<sup>3</sup> and voltage-sensitive calcium channels (VSCCs)<sup>4</sup>, suggesting that CaMKII activation may be initiated within the nanodomains of these channels to produce channel-specific signalling. However, at the level of single synapses it is not known whether CaMKII activation is channel-specific, to what degree CaMKII activation is compartmentalized, and whether it persists in the stimulated spine during LTP.

To measure CaMKII activation in spines, we developed a fluorescence resonance energy transfer (FRET)-based CaMKII $\alpha$  sensor by modifying the previously reported CaMKII $\alpha$  sensor, Camui $\alpha$ , in which the amino and carboxy termini of CaMKII $\alpha$  are labelled with donor and acceptor fluorophores<sup>5</sup>. To optimize the sensitivity and brightness, we used the FRET pair of monomeric enhanced green fluorescent protein (mEGFP)<sup>6</sup> and resonance energy-accepting chromoprotein (REACH), a non-radiative yellow fluorescent protein variant<sup>7</sup> (green-Camui $\alpha$ ), and measured FRET using two-photon fluorescence lifetime imaging microscopy<sup>8–10</sup>. The activation of green-Camui $\alpha$  should change the conformation of CaMKII $\alpha$  to the open state in which its kinase domain is exposed<sup>1</sup>, thereby decreasing FRET and increasing the fluorescence lifetime of mEGFP (Supplementary Fig. 1). We confirmed that green-Camui $\alpha$  is incorporated into a CaMKII dodecameric holoenzyme, and that the fluorescence lifetime of green-Camui $\alpha$  reports CaMKII activation associated with T286 phosphorylation as well as  $\text{Ca}^{2+}/\text{CaM}$  binding (Supplementary Information). We biologically<sup>11</sup> transfected CA1 pyramidal neurons in hippocampal cultured slices with green-Camui $\alpha$  and mCherry. The expression level of green-Camui $\alpha$  was estimated to be 25–50% of the level of endogenous CaMKII $\alpha$  subunit, whose concentration was estimated at ~10–100  $\mu\text{M}$  (Supplementary Fig. 2 and Supplementary Information).

## CaMKII activation during spine structural plasticity

Using this technique, we imaged CaMKII activation during structural plasticity of dendritic spines, which is considered to be associated with

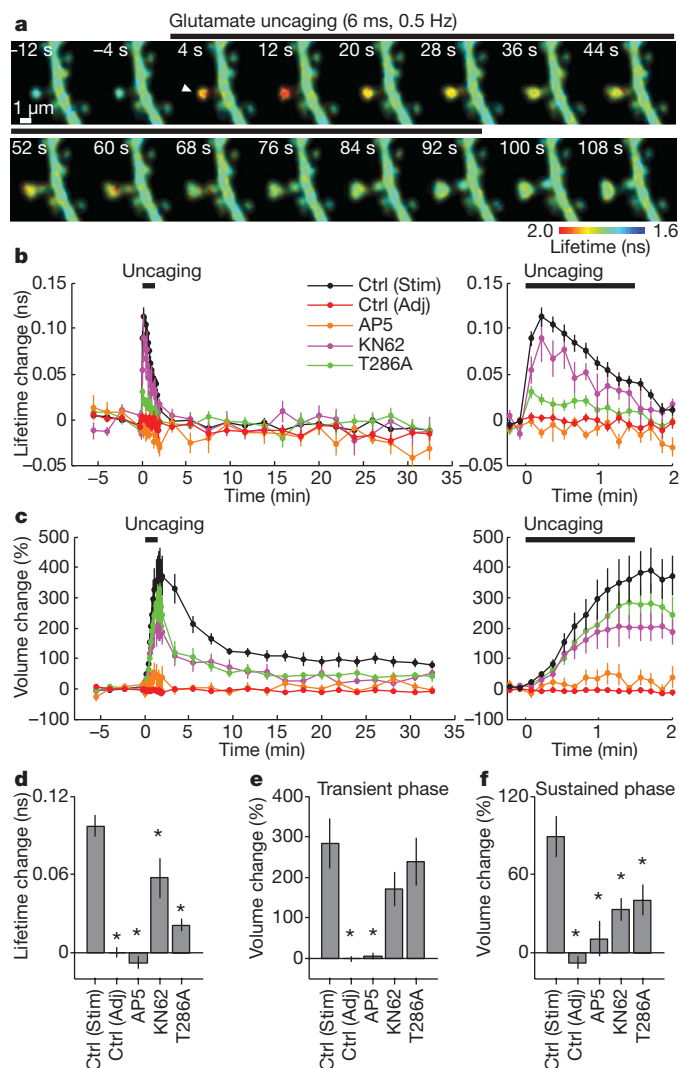
LTP<sup>12–14</sup> (Fig. 1). To induce synapse-specific structural plasticity, we applied a low frequency train of two-photon glutamate uncaging pulses (45 pulses at 0.5 Hz) to a single dendritic spine in zero extracellular  $\text{Mg}^{2+}$  (refs 13, 14). The spine volume increased quickly after glutamate uncaging by  $376 \pm 68\%$  (mean  $\pm$  s.e.m.) (Fig. 1c, e) and relaxed to an elevated level at  $104 \pm 21\%$  for more than 30 min (Fig. 1c, f)<sup>13,14</sup>. The structural plasticity was associated with accumulation of green-Camui $\alpha$  in the stimulated spines that was roughly proportional to the volume change (Supplementary Fig. 3)<sup>15</sup>. Structural plasticity of dendritic spines was abolished by blocking NMDARs (100  $\mu\text{M}$  2-amino-5-phosphonopentanoic acid (AP5))<sup>13</sup> (Fig. 1c, e, f). A CaMKII inhibitor (10  $\mu\text{M}$  KN62) partially inhibited the sustained structural plasticity (Fig. 1c, f), but not the transient phase (Fig. 1c, e), suggesting that the activation of CaMKII is required for the maintenance of the spine volume change<sup>13</sup>. Overexpression of the T286A mutant of green-Camui $\alpha$  reduced the sustained spine enlargement (Fig. 1c, f and Supplementary Fig. 4), demonstrating that this mutant acts as dominant negative. Because the overexpression level of the T286A mutant is relatively small (10–20%; see Supplementary Information), inhibiting autophosphorylation of T286 in a relatively small fraction of CaMKII subunits in a holoenzyme may be sufficient for inhibiting structural plasticity. This result is consistent with previous studies reporting that LTP is attenuated in T286A heterozygous knockout mice<sup>16</sup>.

During structural plasticity, the activity of CaMKII, as measured by the fluorescence lifetime of green-Camui $\alpha$ , increased rapidly (Fig. 1a, b). The activity was restricted to the stimulated spines, and did not spread into dendrites or adjacent spines (Fig. 1a, b, d). Activation of CaMKII preceded the enlargement of spine volume, suggesting that CaMKII triggers molecular processes that lead to the structural plasticity (Fig. 1a, b, c). The NMDAR blocker AP5 inhibited the fluorescence lifetime change, indicating that CaMKII is activated by  $\text{Ca}^{2+}$  from NMDARs (Fig. 1b, d). The CaMKII inhibitor KN62 decreased the fluorescence lifetime change by 40% (Fig. 1b, d). The T286A and T305D mutant sensors reported less activation, suggesting that the binding of calmodulin and the phosphorylation of T286 is required for the full activation of CaMKII in spines (Fig. 1b, d and Supplementary Fig. 4).

## Mechanisms of spine-specific CaMKII activation

In general, the degree of compartmentalization of CaMKII activation is determined by two factors: the inactivation time constant of

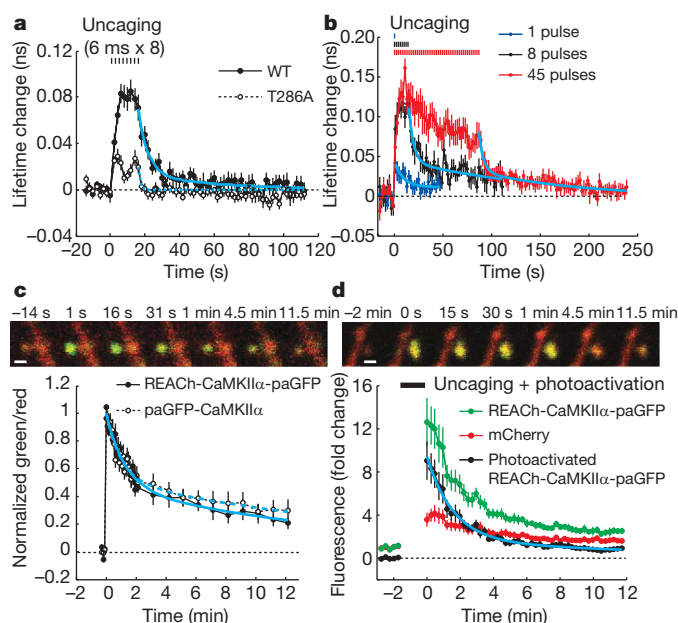
<sup>1</sup>Department of Neurobiology, Duke University Medical Center, Durham, North Carolina 27710, USA.



**Figure 1 | Simultaneous measurements of CaMKII activation and structural plasticity in single spines using two-photon fluorescence lifetime imaging microscopy combined with two-photon glutamate uncaging.**

**a**, Fluorescence lifetime images of green-Camuiα during the induction of spine structural plasticity by two-photon glutamate uncaging in the absence of extracellular  $Mg^{2+}$ . Longer lifetimes indicate increased activity. The white arrowhead indicates the location of uncaging laser spot. **b**, Averaged time course of fluorescence lifetime change of green-Camuiα in the stimulated (Stim) spine and adjacent (Adj) spines (within 5 μm of stimulated spine). Data using pharmacological inhibitors and T286A mutant green-Camuiα are also shown. The number of samples (spines/neurons) is 35/30 for stimulated spines, 29/28 for adjacent spines, 5/4 for AP5, 7/5 for KN62 and 18/11 T286A. Right panel shows expanded view of fluorescence lifetime change during uncaging. **c**, Averaged time course of spine volume change in stimulated spines and adjacent spines. Right panel shows expanded view of the volume change. **d**, CaMKII activation (averaged over 0–40 s or first five points). Asterisks denote statistical significance between control and others ( $P < 0.05$ ). **e**, Transient volume change (average volume change over 1.5–2 min subtracted by that averaged over 25–30 min).<sup>13</sup> **f**, Sustained volume change (volume change averaged over 25–30 min). Error bars are s.e.m.

CaMKII activity, and the mobility of CaMKII (refs 10, 17). Thus, we measured these parameters to study the mechanisms by which CaMKII activation is restricted. To measure the kinetics of CaMKII precisely, we imaged CaMKII activity with higher temporal resolution (2 s) in response to a brief train of glutamate uncaging pulses (0.5 Hz, eight pulses in zero extracellular  $Mg^{2+}$ ). The activation of CaMKII reached a plateau within 6 s, and after the cessation of uncaging, it was inactivated with two time constants, 6 and 45 s



**Figure 2 | Measurements of parameters that determine spine-specificity: inactivation kinetics and mobility of CaMKII.** **a**, Measurements of CaMKII activation and inactivation after brief stimulation (eight pulses, 0.5 Hz) of spines in zero extracellular  $Mg^{2+}$ . The solid cyan curve indicates double exponential functions obtained by fitting to wild-type data (time constants: 6 s (80%) and 45 s (20%)). The dashed cyan curve indicates single exponential function with 1 s time constant. The number of samples (spines/neurons) is 25/5 for wild type, 28/4 for T286A. **b**, CaMKII inactivation kinetics after different number of uncaging pulses (1, 8 and 45 pulses, 0.5 Hz). Data were fitted with double exponential functions (cyan curves). The time constants were 9.3 s (71%) and >100 s (29%) for 1 pulse, 4.7 s (61%) and 136.6 s (39%) for 8 pulses, and 4.5 s (61%) and 102 s (39%) for 45 pulses. The number of samples (spines/neurons) is 28/9 (1 pulses), 16/8 (8 pulses) and 13/7 (45 pulses). **c**, Measurements of spine-dendrite coupling of CaMKII using paGFP. The top panel shows green and red fluorescence before and after photoactivation in a neuron expressing REACH-CaMKIIα-paGFP (green) and mCherry (red). Scale bars, 1 μm. The bottom panel shows the averaged time course of the red/green ratio. Signal was normalized to the peak fluorescence change. The decay time constants were 1.3 min (48%) and 17.0 min (52%) for REACH-CaMKIIα-paGFP (solid cyan curve), and 1.0 min (53%) and 20.7 min (47%) for paGFP-CaMKIIα (dashed cyan curve). The number of samples (spines/neurons) is 15/5 (REACH-CaMKIIα-paGFP) and 22/9 (paGFP-CaMKIIα). **d**, Spine-dendrite coupling of green-Camuiα during structural plasticity. During glutamate uncaging (720 nm, 6 ms, 45 pulses in  $Mg^{2+}$  free solution), REACH-CaMKIIα-paGFP can be photoactivated. We subtracted the normalized red from the normalized green fluorescence, assuming that the small baseline fluorescence of REACH-CaMKIIα-paGFP behaves similarly to the mCherry fluorescence change (see Supplementary Fig. 3). The decay time constants were obtained as 1.6 (82%) and 15.3 (18%) min (cyan curve). The number of samples (spines/neurons) is 10/7. All error bars are s.e.m.

(Fig. 2a). T286A green-Camuiα was activated much less than wild type, and decayed within 2 s (Fig. 2a). These results indicate that T286 phosphorylation prolongs CaMKII activation and allows the accumulation of activated CaMKII during repetitive glutamate uncaging. To measure how the decay kinetics are regulated by CaMKII activation, we stimulated spines with different numbers of uncaging pulses from 1 to 45 (Fig. 2b). With 45 uncaging pulses, we observed enlargements of spines. Our data indicate that the decay time constants depend on neither the number of stimulations nor the level of the CaMKII activation (Fig. 2b).

Next, we measured the mobility of CaMKII, another determinant of the compartmentalization of CaMKII, using photoactivatable GFP (paGFP; Fig. 2c, d). We used either a green-Camuiα variant with EGFP replaced by photoactivatable GFP (REACH-CaMKII-paGFP) or paGFP-tagged CaMKIIα (paGFP-CaMKIIα). After photoactivation



of paGFP, green fluorescence decayed owing to the diffusion of CaMKII $\alpha$  out of the spine with two time constants,  $\sim 1$  and  $\sim 20$  min (Fig. 2c). Fluorescence recovery after photobleaching (FRAP) of green-Camuia $\alpha$  also showed similar recovery time constants (Supplementary Fig. 5b, c). We also measured the mobility of CaMKII during structural plasticity by monitoring the decay of REACH-CaMKII $\alpha$ -paGFP which is photoactivated by glutamate uncaging protocol (6 ms pulse, 45 times). The photoactivated REACH-CaMKII $\alpha$ -paGFP decayed with the time constants similar to the resting condition (Fig. 2d). Consistent with this spine–dendrite coupling time constant, accumulation of CaMKII $\alpha$  into spines during the transient phase of structural plasticity was smaller than the spine volume change (Supplementary Fig. 3). Because the inactivation of CaMKII is faster than its diffusion out of the spine, CaMKII activation must be restricted to the stimulated spine, which agrees with our measurements of CaMKII activity during structural plasticity (Fig. 1).

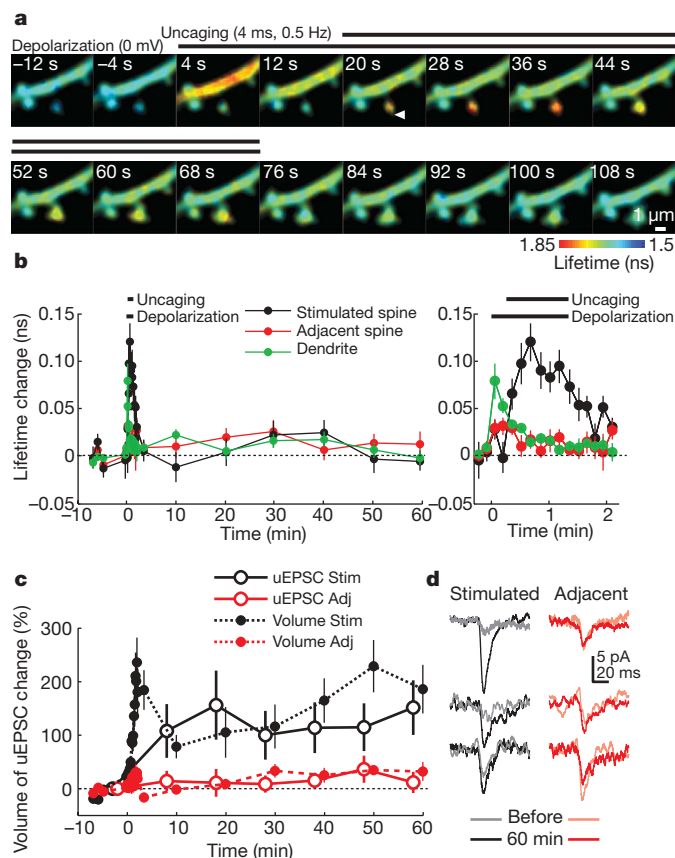
### CaMKII activation during pairing-induced LTP

In the presence of extracellular  $Mg^{2+}$ , LTP induction in hippocampal CA1 synapses requires coincident postsynaptic depolarization and synaptic activation to release  $Mg^{2+}$  block of NMDARs<sup>18</sup>. To study the dynamics of CaMKII activation under such conditions, we imaged CaMKII activation during LTP induced by pairing postsynaptic depolarization with two-photon glutamate uncaging in the presence of extracellular  $Mg^{2+}$  (1 mM)<sup>13</sup> (Fig. 3). We performed whole-cell patch clamp recordings on neurons transfected with green-Camuia $\alpha$  and mCherry, and measured the postsynaptic current evoked by two-photon glutamate uncaging (uncaging-evoked EPSC) at  $-65$  mV. LTP was induced by pairing uncaging pulses (30 pulses at 0.5 Hz) with postsynaptic depolarization (0 mV). This LTP protocol induced a sustained increase in both uncaging-evoked EPSC and spine volume by  $\sim 150\%$ , which was maintained for more than 60 min (Fig. 3c, d). In the same LTP protocol, depolarization produced large calcium increases in spines ( $2.6 \pm 0.5 \mu M$ ) and in dendritic shafts ( $2.6 \pm 0.5 \mu M$ ), presumably due to the activation of VSCCs<sup>19</sup> (Supplementary Fig. 6a–c). Although the depolarization was held for more than 1 min,  $[Ca^{2+}]$  decayed rapidly within 2 s<sup>20</sup> (Supplementary Fig. 6b, c). On top of small residual  $Ca^{2+}$  ( $0.21 \pm 0.07 \mu M$ ), subsequent glutamate uncaging produced calcium transients between  $1.8 \pm 0.3 \mu M$  (first uncaging) and  $0.8 \pm 0.2 \mu M$  (30th uncaging), largely restricted to the heads of the stimulated spines<sup>21,22</sup> (Supplementary Fig. 6b, c).

In response to depolarization, CaMKII activity increased in dendritic shafts (lifetime change =  $0.079 \pm 0.018$  ns), but less so in spines ( $0.028 \pm 0.010$  ns) (Fig. 3a, b). Glutamate uncaging pulses caused further CaMKII activation in the stimulated spines ( $0.097 \pm 0.014$  ns), but not in the adjacent spines (Fig. 3a, b). CaMKII activation in spines and dendritic shafts rapidly decayed within  $\sim 2$  min. The large gradient of CaMKII activation between spines and dendritic shafts during LTP induction demonstrates the high degree of compartmentalization of CaMKII activation in individual spines (Fig. 3a, b).

### Channel-specific activation of CaMKII

It has been reported that  $Ca^{2+}$  through VSCCs is required for LTP induced by several protocols<sup>19,23–26</sup>. Therefore, we further characterized the depolarization-induced CaMKII activation with higher temporal resolution (2 s; Fig. 4). We found that CaMKII activation in dendritic shafts increased within 2 s after depolarization, was sustained during the depolarization (16 s), and then decayed rapidly after repolarization (Fig. 4a). Activation of CaMKII in spines ( $0.034 \pm 0.004$  ns) was smaller than in dendrites ( $0.076 \pm 0.007$  ns).  $[Ca^{2+}]$  during the depolarization decayed within 2 s, and little extra  $Ca^{2+}$  remained for more than 10 s ( $0.21 \pm 0.07 \mu M$ ) (Supplementary Fig. 6b). Thus, these results indicate that small sustained increases in  $[Ca^{2+}]$  are sufficient to sustain CaMKII activation during the depolarization for more than  $\sim 10$  s. The T286A mutant was inactivated much more rapidly (in less than 4 s) than the wild type during the depolarization (Fig. 4b), indicating

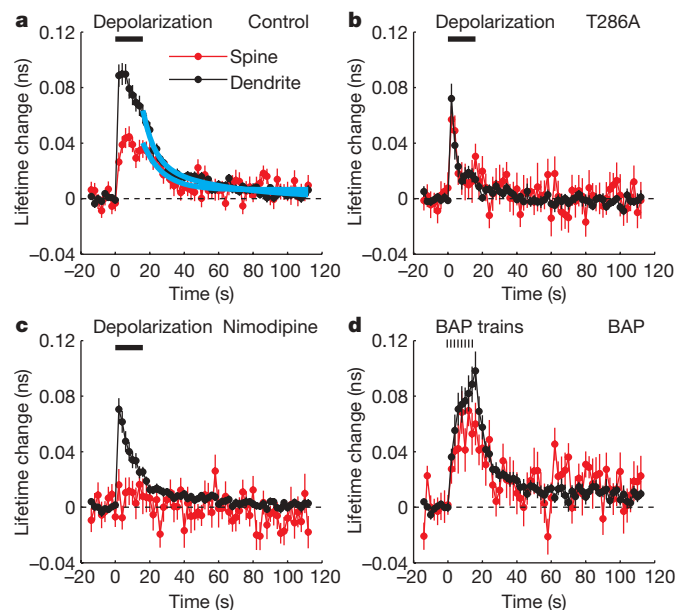


**Figure 3 | Imaging CaMKII activation during LTP induction by two-photon glutamate uncaging paired with postsynaptic depolarization.**

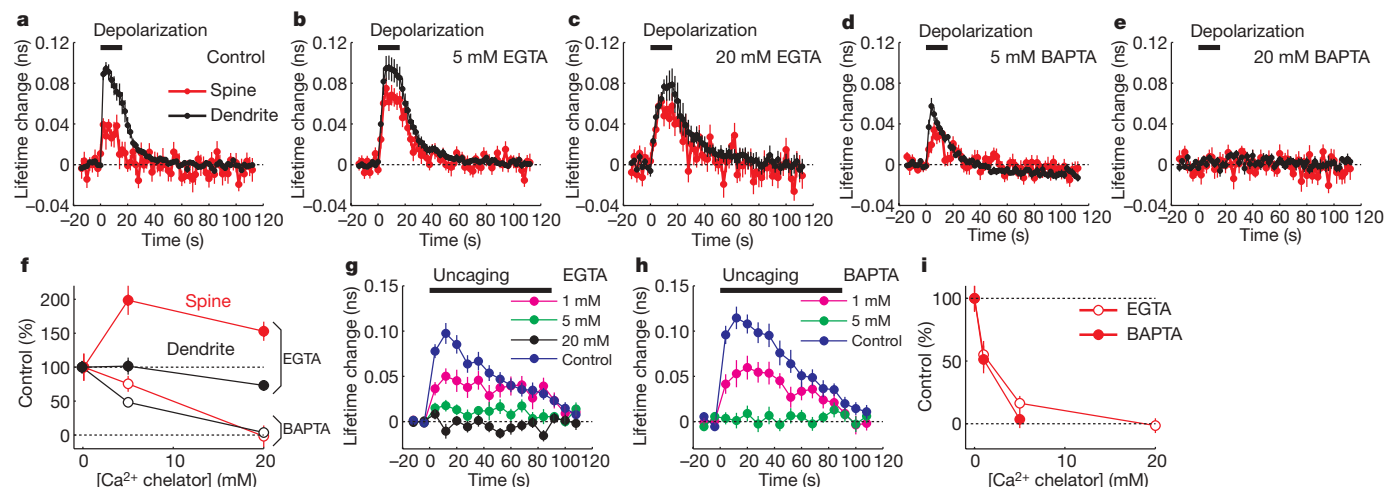
**a**, Fluorescence lifetime images of green-Camuia during the induction of LTP by two-photon glutamate uncaging paired with postsynaptic depolarization. **b**, Averaged time course of fluorescence lifetime change of green-Camuia in the stimulated spine, adjacent spines (within  $5 \mu m$  of stimulated spine), and dendritic shaft close to the stimulated spine. Right panel shows extended view of fluorescence lifetime change during uncaging. Results are the average of 8 stimulated spines, 13 adjacent spines and 8 dendrites from 8 neurons. The plot of stimulated spines is noisier due to the smaller number of samples. Plots of stimulated spines in **b–d** are from the same spines. **c**, Averaged time course of changes in spine volume and the amplitude of uncaging-evoked EPSC (uEPSC) at  $-65$  mV in the stimulated (Stim) spine and adjacent (Adj) spines (within  $5 \mu m$ ). For uncaging-evoked EPSC, each point is the average of four trials at 5-s intervals. The uncaging spot was relocated to the tip of spines before each of the four trial measurements. **d**, Uncaging-evoked EPSC at single spines before and 60 min after LTP induction (average of four trials, filtered with 2-ms window). All error bars are s.e.m.

that the prolonged activation of CaMKII requires autophosphorylation at T286, and the decay of CaMKII activity after repolarization is due to the dephosphorylation of T286. Bursts of back-propagating action potentials (83 Hz, 0.5 s, eight times, 2 s interval) also produced similar CaMKII activation in spines and dendrites (Fig. 4d).

Nimodipine, a blocker of L-type VSCCs, abolished CaMKII activation in dendritic spines during depolarization (Fig. 4c). The peak of CaMKII activation in dendritic shafts was not affected by the blockade of L-type VSCCs, but the activity decayed more rapidly (Fig. 4c). Because L-type VSCCs produce only a small fraction of  $Ca^{2+}$  in spines and dendritic shafts during cell depolarization (Supplementary Fig. 6d), L-type VSCC-dependent CaMKII activation presumably occurs in the nanodomain at the mouth of L-type VSCCs<sup>4,27,28</sup>. Also, global  $Ca^{2+}$  increase by opening of non-L-type VSCCs did not activate CaMKII in spines (Fig. 4c), although it is as large as uncaging-evoked  $Ca^{2+}$  increase (Supplementary Fig. 6). This result indicates that global  $Ca^{2+}$  increase alone is insufficient to activate CaMKII in spines.



**Figure 4 | Differential activation of CaMKII in spines and dendrites by postsynaptic depolarization.** **a**, High temporal resolution measurements of the fluorescence lifetime change of green-Camuix in spines and dendrites in response to postsynaptic depolarization (0 mV, 16 s, indicated by the bar) in spines (red) and dendritic shafts (black). Results are the average of 53 spines and 30 dendrites from 12 neurons. The cyan lines are a double exponential function obtained by fitting. The time constants are 11.3 s (84%) and >300 s (16%) for spines, and 6.8 s (60%) and 43.7 s (40%) for dendrites. **b**, Fluorescence lifetime change of T286A mutant of green-Camuix. Results are the average of 19 spines and 14 dendrites from 8 neurons. **c**, Fluorescence lifetime change of green-Camuix in the presence of L-type VSCC inhibitor (20  $\mu$ M Nimodipine). Results are the average of 22 spines and 15 dendrites from 7 neurons. **d**, Bursts of back-propagating action potentials (BAP)-induced CaMKII activation at proximal dendrites (<100  $\mu$ m). We applied eight back-propagating action potential bursts, each 83 Hz for 0.5 s, at 2-s intervals. Results are the average of 22 spines and 12 dendrites from 7 neurons. All error bars are s.e.m.



**Figure 5 | Effects of  $\text{Ca}^{2+}$  chelators on CaMKII activation.** **a–e**, Fluorescence lifetime change of green-Camuix in spines and dendrites in response to postsynaptic depolarization (0 mV, 0–16 s, indicated by the black bar) in spines (red) and dendritic shafts (black) in neurons patch-clamped with electrodes containing no  $\text{Ca}^{2+}$  chelator (**a**), 5 mM EGTA (**b**), 20 mM EGTA (**c**), 5 mM BAPTA (**d**) or 20 mM BAPTA (**e**). The number of samples (spines/dendrites/neurons) is 18/11/6 for control (**a**), 29/14/7 for 5 mM EGTA (**b**), 15/7/4 for 20 mM EGTA (**c**), 21/7/5 for 5 mM BAPTA (**d**) and 18/7/4 for 20 mM BAPTA (**e**). Imaging was performed more than 15 min after establishing a whole-cell patch. **f**, Dependence of fluorescence lifetime change (averaged over 0–16 s) by depolarization on  $\text{Ca}^{2+}$  chelator concentration (EGTA: closed circle, BAPTA: open circle, red: spine, black: dendrite). Fluorescence lifetime was normalized to control. **g, h**, Fluorescence

To test further the hypothesis of the nanodomain activation of CaMKII, we performed whole-cell patch clamping using a pipette including the  $\text{Ca}^{2+}$  chelators EGTA or BAPTA. BAPTA and EGTA have similar dissociation constants to  $\text{Ca}^{2+}$ , but the binding rate of BAPTA is  $\sim 100$  times higher<sup>29</sup>. Thus, EGTA inhibits global  $\text{Ca}^{2+}$  over nanodomain  $\text{Ca}^{2+}$  more selectively than BAPTA does<sup>29</sup>. The activation of CaMKII induced by depolarization was inhibited by BAPTA much more effectively compared to the same concentration of EGTA (Fig. 5a–f), suggesting that nanodomain  $\text{Ca}^{2+}$  near calcium channels is sufficient to activate CaMKII both in spines and in dendrites. With EGTA loaded in neurons, depolarization induced larger CaMKII activation in spines, suggesting that there is global  $\text{Ca}^{2+}$ -dependent CaMKII inactivation (Fig. 5b, c). Slower CaMKII activation kinetics in the presence of EGTA (Fig. 5b, c) are presumably caused by smaller  $\text{Ca}^{2+}$ -dependent inactivation combined with smaller activation per unit time due to reduced  $\text{Ca}^{2+}$  increase. Unlike depolarization-induced CaMKII activation, EGTA and BAPTA both similarly reduced glutamate-uncaging-induced CaMKII activation (Fig. 5g–i), suggesting that nanodomain  $\text{Ca}^{2+}$  alone is not sufficient for uncaging-induced CaMKII activation.

## Discussion

In this study, we visualized CaMKII activation in single dendritic spines during LTP. Using this method, we demonstrated that CaMKII activation is largely restricted to the stimulated spine during LTP induced by two-photon glutamate uncaging (Figs 1 and 3). This suggests that CaMKII activation is important for the synapse specificity of LTP. This finding is in contrast to the activation of Ras, which is another molecule essential for LTP induction<sup>30</sup>. In response to LTP-inducing stimuli, Ras is activated in stimulated spines, and then activated Ras spreads over  $\sim 10 \mu\text{m}$  of dendritic length and invades nearby spines by diffusion<sup>10</sup>. The restricted CaMKII activation can be explained by the rapid CaMKII inactivation in spines and dendrites ( $\sim 5$  s and  $\sim 1$  min, Figs 2a, b and 4a) and slow spine-dendrite diffusion coupling ( $\sim 1$  and  $\sim 20$  min, Fig. 2c, d)<sup>31</sup>: activated CaMKII at the stimulated spine gets

lifetime change of green-Camuix in response to glutamate uncaging (6 ms, 45 stimuli in zero extracellular  $\text{Mg}^{2+}$ ) in neurons patch-clamped with electrodes containing 1–20 mM EGTA (**g**) or 1–5 mM BAPTA (**h**) in a current-clamp mode. Controls are measured in the same neurons before patch-clamping. Patch-clamping with internal solution without  $\text{Ca}^{2+}$  chelator did not alter the fluorescence lifetime change (Supplementary Fig. 7). The number of samples (spines/neurons) is 22/9 for control in **g**, 12/3 for 1 mM EGTA, 15/3 for 5 mM EGTA, 17/3 for 20 mM EGTA, 16/7 for control in **h**, 17/4 for 1 mM BAPTA and 14/3 for 5 mM BAPTA. **i**, Dependence of fluorescence lifetime change (averaged over 0–40 s) induced by glutamate uncaging on  $\text{Ca}^{2+}$  chelator concentration (EGTA: open circle, BAPTA: closed circle). Fluorescence lifetime was normalized to control. Error bars represent s.e.m.



inactivated before it diffuses out of the spine. In contrast, Ras inactivation is much slower ( $\sim 5$  min) and the spine–dendrite coupling time constant is smaller ( $\sim 5$  s)<sup>10</sup>. This allows activated Ras to spread and invade nearby spines before it gets inactivated. Because CaMKII and Ras act in parallel pathways to induce LTP<sup>10</sup>, the spine specificity of LTP may be produced by CaMKII, whereas Ras may be permissive for LTP and responsible for dendritic level signalling such as the regulation of threshold for LTP induction<sup>10,14</sup>.

Rapid inactivation of CaMKII during LTP induction (Figs 1 and 3) is consistent with previous studies showing that persistent activation of CaMKII is not required for LTP maintenance<sup>32,33</sup>. However, this seems to be inconsistent with biochemical studies demonstrating that T286 phosphorylation is sustained after LTP induction<sup>34,35</sup>. This discrepancy may be due to the difference in sample preparation (acute slices versus cultured slices), in the population of CaMKII measured (whole lysate including soma versus single synapses), or in stimulation models (most of the previous biochemical studies used strong stimulations to induce LTP in a large number of synapses, whereas we induced LTP in a single synapse). Alternatively, only the subpopulation of CaMKII subunits that bind to NR2B subunit of NMDARs may maintain their activity<sup>3,36</sup>. Given that there is more CaMKII than NMDARs in the postsynaptic density<sup>37</sup>, the subpopulation of CaMKII that undergoes persistent activation may be below the detection threshold of our method.

We found that T286 phosphorylation is required for high CaMKII activation in response to repetitive two-photon uncaging (Figs 1b and 2a).  $[Ca^{2+}]$  transients last only  $\sim 0.2$  s after each uncaging (Supplementary Fig. 6a–c), owing to strong  $Ca^{2+}$  extrusion in spines and dendrites<sup>17</sup>. Thus,  $Ca^{2+}$ /CaM should dissociate from CaMKII rapidly after these stimulations as indicated in the activation time course of T286A green-Camuia (Figs 2a and 4b). The phosphorylation of T286 prolongs the activation of CaMKII by  $\sim 5$  s (Figs 2a and 4a), and allows for the accumulation of activated CaMKII during repetitive uncaging (Fig. 2a).

Our imaging results indicate that CaMKII activation in dendritic spines and dendritic shafts are regulated by different  $Ca^{2+}$  sources in response to cell depolarization: CaMKII in dendritic spines is activated specifically by L-type VSCCs, whereas CaMKII in dendritic shafts is activated mostly by non-L-type channels in addition to a small contribution from L-type VSCCs (Fig. 4). Because L-type VSCCs do not contribute to the global  $Ca^{2+}$  in spines and dendritic shafts (Supplementary Fig. 6d), the L-type-VSCC-dependent activation of CaMKII must be produced in the nanodomain at the mouth of L-type VSCCs<sup>4,27,28</sup>. Experiments using  $Ca^{2+}$  chelators with different binding rates (Fig. 5a–f) also suggest that nanodomain  $Ca^{2+}$  at calcium channels is sufficient for CaMKII activation in response to cell depolarization, both in spines and dendrites. Because L-type VSCCs are required for hippocampal LTP induced by several protocols<sup>19,23–26</sup>, CaMKII activation by L-type VSCCs, although smaller than that by NMDARs (Figs 3 and 4), is probably important for the modulation of LTP.

Unlike CaMKII activation in response to depolarization, nanodomain  $Ca^{2+}$  alone was insufficient for NMDAR-mediated CaMKII activation (Fig. 5g–i). Because global  $Ca^{2+}$  increase through non-L-type VSCCs was not sufficient to activate CaMKII in spines (Fig. 4c) (although it is as high as and longer-lasting than the uncaging-evoked  $[Ca^{2+}]$  increase; Supplementary Fig. 6d), NMDAR-mediated CaMKII activation presumably requires both global and nanodomain  $Ca^{2+}$ . Thus, CaMKII activation mechanisms depend on the channels from which  $Ca^{2+}$  influx occurs. The difference may be due to different CaM localization near channels. Also, the dynamics of  $Ca^{2+}$ -dependent protein phosphatase activity are probably important in shaping the spatiotemporal dynamics of CaMKII (Fig. 5a–f).

Most LTP induction protocols involve postsynaptic depolarization<sup>18</sup>, which alone can produce large  $Ca^{2+}$  in spines (Supplementary Fig. 6)<sup>19</sup>. However, it had remained a mystery as to why postsynaptic depolarization alone cannot induce synapse nonspecific

LTP<sup>18,19</sup>. Our results demonstrate that depolarization activates only a subset of CaMKII that interact with L-type VSCCs, which presumably is insufficient for LTP induction<sup>38</sup>. Thus, local  $Ca^{2+}$ -mediated, channel-specific CaMKII activation in spines has an important role in the induction of synapse-specific LTP.

## METHODS SUMMARY

Green-Camuia was made from Camuia<sup>5</sup> by replacing ECFP with mEGFP and Venus with REACh (Venus(A206K/T145W)<sup>7,39</sup>). Hippocampal cultured slices were prepared from postnatal day 6–8 rats as described<sup>40</sup>. Neurons were sparsely transfected with green-Camuia and mCherry using ballistic gene transfer<sup>11</sup> at days 8–14 *in vitro*, and imaged 2–5 days after transfection. CaMKII activity was measured using two-photon fluorescence lifetime imaging microscopy and spine volume change was monitored by measuring the fluorescence intensity of mCherry (red) in spines using regular two-photon microscopy<sup>9,10</sup>. Statistical tests were performed using analysis of variance (ANOVA) followed by the least significant difference post-hoc tests. All error bars indicate s.e.m.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 9 November 2008; accepted 29 January 2009.

- Rosenberg, O. S. *et al.* Oligomerization states of the association domain and the holoenzyme of  $Ca^{2+}$ /CaM kinase II. *FEBS J.* **273**, 682–694 (2006).
- Lisman, J. E. & Zhabotinsky, A. M. A model of synaptic memory: a CaMKII/PP1 switch that potentiates transmission by organizing an AMPA receptor anchoring assembly. *Neuron* **31**, 191–201 (2001).
- Merrill, M. A., Chen, Y., Strack, S. & Hell, J. W. Activity-driven postsynaptic translocation of CaMKII. *Trends Pharmacol. Sci.* **26**, 645–653 (2005).
- Hudmon, A. *et al.* CaMKII tethers to L-type  $Ca^{2+}$  channels, establishing a local and dedicated integrator of  $Ca^{2+}$  signals for facilitation. *J. Cell Biol.* **171**, 537–547 (2005).
- Takao, K. *et al.* Visualization of synaptic  $Ca^{2+}$ /calmodulin-dependent protein kinase II activity in living neurons. *J. Neurosci.* **25**, 3107–3112 (2005).
- Zacharias, D. A., Violin, J. D., Newton, A. C. & Tsien, R. Y. Partitioning of lipid-modified monomeric GFPs into membrane microdomains of live cells. *Science* **296**, 913–916 (2002).
- Ganesan, S., Ameer-Beg, S. M., Ng, T. T., Vojnovic, B. & Wouters, F. S. A dark yellow fluorescent protein (YFP)-based Resonance Energy-Accepting Chromoprotein (REACh) for Förster resonance energy transfer with GFP. *Proc. Natl Acad. Sci. USA* **103**, 4089–4094 (2006).
- Yasuda, R. Imaging spatiotemporal dynamics of neuronal signaling using fluorescence resonance energy transfer and fluorescence lifetime imaging microscopy. *Curr. Opin. Neurobiol.* **16**, 551–561 (2006).
- Yasuda, R. *et al.* Super-sensitive Ras activation in dendrites and spines revealed by 2-photon fluorescence lifetime imaging. *Nature Neurosci.* **9**, 283–291 (2006).
- Harvey, C. D., Yasuda, R., Zhong, H. & Svoboda, K. The spread of Ras activity triggered by activation of a single dendritic spine. *Science* **321**, 136–140 (2008).
- McAllister, A. K. Biolistic transfection of neurons. *Sci. STKE* **2000**, pl1 (2000).
- Matsuzaki, M. *et al.* Dendritic spine geometry is critical for AMPA receptor expression in hippocampal CA1 pyramidal neurons. *Nature Neurosci.* **4**, 1086–1092 (2001).
- Matsuzaki, M., Honkura, N., Ellis-Davies, G. C. & Kasai, H. Structural basis of long-term potentiation in single dendritic spines. *Nature* **429**, 761–766 (2004).
- Harvey, C. D. & Svoboda, K. Locally dynamic synaptic learning rules in pyramidal neuron dendrites. *Nature* **450**, 1195–1200 (2007).
- Otmakhov, N. *et al.* Persistent accumulation of calcium/calmodulin-dependent protein kinase II in dendritic spines after induction of NMDA receptor-dependent chemical long-term potentiation. *J. Neurosci.* **24**, 9324–9331 (2004).
- Ohno, M., Frankland, P. W. & Silva, A. J. A pharmacogenetic inducible approach to the study of NMDA/ $\alpha$ CaMKII signaling in synaptic plasticity. *Curr. Biol.* **12**, 654–656 (2002).
- Sabatini, B. L., Oertner, T. G. & Svoboda, K. The life cycle of  $Ca^{2+}$  ions in dendritic spines. *Neuron* **33**, 439–452 (2002).
- Malenka, R. C. & Bear, M. F. LTP and LTD: an embarrassment of riches. *Neuron* **44**, 5–21 (2004).
- Conti, R. & Lisman, J. A large sustained  $Ca^{2+}$  elevation occurs in unstimulated spines during the LTP pairing protocol but does not change synaptic strength. *Hippocampus* **12**, 667–679 (2002).
- Magee, J. C. & Johnston, D. Characterization of single voltage-gated  $Na^{+}$  and  $Ca^{2+}$  channels in apical dendrites of rat CA1 pyramidal neurons. *J. Physiol. (Lond.)* **487**, 67–90 (1995).
- Noguchi, J., Matsuzaki, M., Ellis-Davies, G. C. & Kasai, H. Spine-neck geometry determines NMDA receptor-dependent  $Ca^{2+}$  signaling in dendrites. *Neuron* **46**, 609–622 (2005).
- Sobczyk, A., Scheuss, V. & Svoboda, K. NMDA receptor subunit-dependent  $[Ca^{2+}]$  signaling in individual hippocampal dendritic spines. *J. Neurosci.* **25**, 6037–6046 (2005).

23. Dan, Y. & Poo, M. M. Spike timing-dependent plasticity of neural circuits. *Neuron* **44**, 23–30 (2004).
24. Grover, L. M. & Teyler, T. J. Two components of long-term potentiation induced by different patterns of afferent activation. *Nature* **347**, 477–479 (1990).
25. Magee, J. C. & Johnston, D. A synaptically controlled, associative signal for Hebbian plasticity in hippocampal neurons. *Science* **275**, 209–213 (1997).
26. Remy, S. & Spruston, N. Dendritic spikes induce single-burst long-term potentiation. *Proc. Natl Acad. Sci. USA* **104**, 17192–17197 (2007).
27. Yasuda, R., Sabatini, B. L. & Svoboda, K. Plasticity of calcium channels in dendritic spines. *Nature Neurosci.* **6**, 948–955 (2003).
28. Wheeler, D. G., Barrett, C. F., Groth, R. D., Safa, P. & Tsien, R. W. CaMKII locally encodes L-type channel activity to signal to nuclear CREB in excitation-transcription coupling. *J. Cell Biol.* **183**, 849–863 (2008).
29. Neher, E. Usefulness and limitations of linear approximations to the understanding of  $\text{Ca}^{2+}$  signals. *Cell Calcium* **24**, 345–357 (1998).
30. Zhu, J. J., Qin, Y., Zhao, M., Van Aelst, L. & Malinow, R. Ras and Rap control AMPA receptor trafficking during synaptic plasticity. *Cell* **110**, 443–455 (2002).
31. Okamoto, K., Narayanan, R., Lee, S. H., Murata, K. & Hayashi, Y. The role of CaMKII as an F-actin-bundling protein crucial for maintenance of dendritic spine structure. *Proc. Natl Acad. Sci. USA* **104**, 6418–6423 (2007).
32. Malinow, R., Schulman, H. & Tsien, R. W. Inhibition of postsynaptic PKC or CaMKII blocks induction but not expression of LTP. *Science* **245**, 862–866 (1989).
33. Chen, H. X., Otmakhov, N., Strack, S., Colbran, R. J. & Lisman, J. E. Is persistent activity of calcium/calmodulin-dependent kinase required for the maintenance of LTP? *J. Neurophysiol.* **85**, 1368–1376 (2001).
34. Fukunaga, K., Muller, D. & Miyamoto, E. Increased phosphorylation of  $\text{Ca}^{2+}$ /calmodulin-dependent protein kinase II and its endogenous substrates in the induction of long-term potentiation. *J. Biol. Chem.* **270**, 6119–6124 (1995).
35. Barria, A., Muller, D., Derkach, V., Griffith, L. C. & Soderling, T. R. Regulatory phosphorylation of AMPA-type glutamate receptors by CaM-KII during long-term potentiation. *Science* **276**, 2042–2045 (1997).
36. Lisman, J., Schulman, H. & Cline, H. The molecular basis of CaMKII function in synaptic and behavioural memory. *Nature Rev. Neurosci.* **3**, 175–190 (2002).
37. Cheng, D. *et al.* Relative and absolute quantification of postsynaptic density proteome isolated from rat forebrain and cerebellum. *Mol. Cell. Proteomics* **5**, 1158–1170 (2006).
38. Kullmann, D. M., Perkel, D. J., Manabe, T. & Nicoll, R. A.  $\text{Ca}^{2+}$  entry via postsynaptic voltage-sensitive  $\text{Ca}^{2+}$  channels can transiently potentiate excitatory synaptic transmission in the hippocampus. *Neuron* **9**, 1175–1183 (1992).
39. Murakoshi, H., Lee, S.-J. & Yasuda, R. Highly sensitive and quantitative FRET-FLIM imaging in single dendritic spines using improved non-radiative YFP. *Brain Cell Biol.* **36**, 31–42 (2008).
40. Stoppini, L., Buchs, P. A. & Muller, D. A simple method for organotypic cultures of nervous tissue. *J. Neurosci. Methods* **37**, 173–182 (1991).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank H. Murakoshi for purified mEGFP and technical advice, Y. Hayashi for complementary DNAs and shRNA, A. Wang for cultured slices, T. Zimmerman and D. Kloetzer for laboratory management, and M. Ehlers, C. Harvey, J. Lisman, M. Patterson, S. Raghavachari, K. Svoboda, R. Weinberg and H. Zhong for comments on the manuscript. This study was funded by Burroughs Wellcome Fund, Dana Foundation, Alfred P. Sloan Foundation, Autism Speaks, National Alliance of Autism Research, Whitehall Foundation, Alzheimer's Association, National Institute of Mental Health (R01MH08004), National Science Foundation (0642000), Duke University Undergraduate Research Support Grants (S.-J.R.L.), Howard Hughes Neuroscience Forum Fellowship (S.-J.R.L.), and Ruth K. Broad Biomedical Research Foundation (E.M.S.).

**Author Contributions** S.-J.R.L. and R.Y. designed the experiments. R.Y. built the microscope and developed the green-Camuia construct. S.-J.R.L. performed most of the experiments. S.-J.R.L. and Y.E.-R. performed calcium imaging. E.M.S. performed immunostaining and pull-down assay. S.-J.R.L. and R.Y. analysed the data. R.Y. wrote the paper. All authors discussed and commented on the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to R.Y. ([yasuda@neuro.duke.edu](mailto:yasuda@neuro.duke.edu)).



## METHODS

**Constructs.** Camuix (Venus-CaMKII $\alpha$ -CFP) and short-hairpin RNA (shRNA) against CaMKII $\alpha$  was provided by Y. Hayashi<sup>5</sup>. Green-Camuix was made by exchanging Venus with REACH (Venus<sup>11</sup> with monomeric mutation, A206K (ref. 6), and dark mutation, T145W (refs 7, 39, 42)) and CFP with mEGFP (EGFP(A206K))<sup>39</sup>.

**Preparation.** Hippocampal slice cultures were prepared from postnatal day 6 or 7 rats, as described<sup>40</sup>. After 9–14 days in culture, cells were transfected with ballistic gene transfer<sup>11</sup> using gold beads (9–11 mg) coated with plasmids containing cDNA of green-Camuix (35  $\mu$ g) and mCherry (16  $\mu$ g). We used one neuron per slice for most of our experiments. Samples were prepared in accordance with the animal care and use guidelines of Duke University Medical Center.

**Imaging.** Details of FRET imaging using two-photon fluorescence lifetime imaging microscopy have been described previously<sup>9,39</sup>. Green-Camuix and mCherry were simultaneously excited with a Ti:Sapphire laser tuned at 900 or 920 nm<sup>39</sup>. Both epi- and trans-fluorescence were used for the detection. The fluorescence was divided with a dichroic mirror (565 nm) and detected with photoelectron multiplier tubes (PMTs) placed after wavelength filters (Chroma, 510 nm/70 nm (centre wavelength/bandwidth) for green and 620 nm/90 nm for red). For fluorescence lifetime imaging in the green channel, PMT with low transit time spread (Hamamatsu) placed at epi-fluorescence site was used. For intensity imaging, wide aperture PMTs were used (Hamamatsu). Fluorescence lifetime image was produced on a PCI board (Becker-Hickl) controlled with custom software written using Matlab 7.0.

**Fluorescence lifetime image analysis.** Fluorescence lifetime of green-Camuix should have at least three components including open conformation, close conformation, and donor with unfolded acceptor<sup>9</sup>. Thus, it is difficult to make a complete model and perform curve fitting. Instead, the mean fluorescence lifetime averaged over several populations  $\tau_m$  was measured from the mean photon arrival time  $\langle t \rangle$  as follows<sup>9</sup>:

$$\tau_m = \langle t \rangle - t_0 = \frac{\int dt \cdot tF(t)}{\int dt \cdot F(t)} - t_0$$

in which  $F(t)$  is the fluorescence lifetime decay curve (Supplementary Fig. 1a),  $t$  is time, and  $t_0$  is the offset. We estimated the offset  $t_0$  before each experiment by fitting fluorescence lifetime decay curve of green-Camuix expressed in a neuron with double exponential function and comparing  $\langle t \rangle$  with the fluorescence lifetime averaged over two populations as follows:

$$t_0 = \frac{\int dt \cdot tF(t)}{\int dt \cdot F(t)} - \tau_m \sim \frac{\int dt \cdot tF(t)}{\int dt \cdot F(t)} - \frac{P_1\tau_1^2 + P_2\tau_2^2}{P_1\tau_1 + P_2\tau_2}$$

in which  $P_1$  and  $P_2$  are the fraction and  $\tau_1$  and  $\tau_2$  are the fluorescence lifetime of the two populations obtained by fitting. Change in fluorescence lifetime is independent of  $t_0$ , therefore it depends on neither model nor curve fitting.

The fluorescence lifetime of green-Camuix in spines under resting condition was  $1.82 \pm 0.02$  ns for wild type (samples in Fig. 1 and Supplementary Fig. 4 pooled together),  $1.84 \pm 0.02$  ns for T286A (Fig. 1 and Supplementary Fig. 4) and  $1.81 \pm 0.03$  ns for T305D (Supplementary Fig. 4). The fluorescence lifetime under resting condition in the presence of pharmacological drugs was  $1.77 \pm 0.06$  ns for AP5 (Fig. 1),  $1.67 \pm 0.03$  ns for KN62 (Fig. 1),  $1.81 \pm 0.02$  ns for Nimodipine (Fig. 4),  $1.80 \pm 0.01$  ns for 1–20 mM EGTA (Fig. 5), and  $1.83 \pm 0.02$  ns for 1–20 mM BAPTA (Fig. 5). Thus, most of the drugs used in this study did not change the fluorescence lifetime, except that KN62 decreased the fluorescence lifetime significantly ( $P < 0.05$ , comparison between control and KN62 for samples in Fig. 1).

**Two-photon glutamate uncaging.** A laser tuned at 720 nm was used to uncage 4-methoxy-7-nitroindolyl (MNI)-caged L-glutamate (2 mM) in extracellular solution. The same galvanometer-scanning mirrors were used to steer both imaging and uncaging lasers. The laser spot was steered to the tip of a spine image (Supplementary Fig. 8), and 4–6 ms pulses at 4 mW were applied to stimulate glutamate receptors on the spine. This produced a 2–20 pA postsynaptic current (Fig. 3d, Supplementary Fig. 8). The resolution of glutamate uncaging was  $\sim 1 \mu$ m (Supplementary Fig. 8). The intensity of the laser was modulated by a pockels cell (Con-Optics). For LTP experiment (Fig. 3), the uncaging spot was relocated before each uncaging session (each session includes four trials of glutamate uncaging at 0.2 Hz). The session was repeated every 10 min (Fig. 3c). Series and input resistance were continuously monitored, and if they changed more than 20%, the measurements were rejected from further analysis. To induce structural plasticity of spines (Figs 1, 2a, b, d and 5g–i), we used ACSF (130 mM NaCl, 2.5 mM KCl, 2 mM NaHCO<sub>3</sub>, 1.25 mM NaH<sub>2</sub>PO<sub>4</sub> and 25 mM glucose) containing 4 mM CaCl<sub>2</sub>, 1  $\mu$ M tetrodotoxin (TTX) and 2 mM MNI-caged L-glutamate at 25–27 °C. To induce LTP with the pairing protocol (Fig. 3), we added 1 mM MgCl<sub>2</sub>.

**FRAP (and photoactivation).** For FRAP (Supplementary Fig. 4b, c), spine fluorescence was photobleached by parking the imaging laser (tuned at 920 nm) for 1–8 s at a spine of interest. To activate paGFP (Fig. 2c, d), the uncaging laser (tuned at 800 nm) was parked for 20 ms at a spine of interest.

**Electrophysiology.** Whole-cell patch clamping was performed with patch pipettes (4–7 M $\Omega$ ). We used cells with series resistance lower than 25 M $\Omega$  for experiments. To induce LTP with the pairing protocol (Fig. 3) or to induce cell depolarization (Figs 4a–d and 5a–f), we performed experiments in a voltage-clamp mode using Cs<sup>+</sup> internal solution (130 mM CsMeSO<sub>3</sub>, 10 mM sodium phosphocreatine, 4 mM MgCl<sub>2</sub>, 4 mM Na<sub>2</sub>ATP, 0.4 mM Na<sub>2</sub>GTP, 10 mM Cs-HEPES, pH 7.3). To induce back-propagating action potentials (Fig. 4d), we injected 4-nA, 4-ms current pulses in a current-clamp mode using K<sup>+</sup> based internal solution (Cs<sup>+</sup> was substituted with K<sup>+</sup>) and ACSF with 4 mM CaCl<sub>2</sub>, 4 mM MgCl<sub>2</sub>, 1  $\mu$ M TTX, 100  $\mu$ M AP5 and 10  $\mu$ M 2,3-dioxo-6-nitro-1,2,3,4-tetrahydrobenzo[f]quinoxaline-7-sulphonamide (NBQX). For experiments using BAPTA or EGTA (Fig. 5), we started imaging  $\sim 15$  min after establishing the whole-cell patch clamp with pipette including BAPTA or EGTA. Depolarization (Fig. 5a–f) and glutamate uncaging (Fig. 5g–i) were performed in voltage and current-clamp modes, respectively. To prepare internal solution with 1, 5 and 20 mM EGTA or BAPTA (Fig. 5), we prepared 100 mM K-EGTA and K-BAPTA (pH 7.3), and mixed 1%, 5% and 20% volume of this solution and 99%, 95% and 80% volume of Cs<sup>+</sup> (Fig. 5a–f) or K<sup>+</sup> (Fig. 5g–i) based internal solution, respectively. Then, we re-adjusted the osmolarity to 290–300 and the pH to 7.3.

**Calcium imaging.** Ca<sup>2+</sup> imaging (Supplementary Fig. 6) was performed as described previously<sup>43</sup>. We performed whole-cell patch clamping to a neuron in a cultured hippocampal slice using an electrode containing Ca<sup>2+</sup>-sensitive green dye (500  $\mu$ M Fluo-4FF) and Ca<sup>2+</sup>-insensitive red dye (300  $\mu$ M Alexa-594) in Cs<sup>+</sup> internal solution. Images were acquired every 32 ms (Supplementary Fig. 6a–c) or 256 ms (Supplementary Fig. 6d). The ratio of the fluorescence change of green normalized to red signal averaged over an imaging session ( $\Delta G/R$ ) was used as a measure of the [Ca<sup>2+</sup>] transient. Both fluorophores were simultaneously excited with a Ti:Sapphire laser tuned at 920 nm. The change in [Ca<sup>2+</sup>] ( $\Delta[Ca^{2+}]$ ) was measured as:

$$\Delta[Ca^{2+}] = \frac{(\Delta G/R) K_D}{(G/R)_{sat}}$$

in which  $(G/R)_{sat}$  is the ratio of green and red signals at saturating Ca<sup>2+</sup> (10 mM) measured in a pipette, and the dissociation constant ( $K_D$ ) is 10.4  $\mu$ M for Fluo-4FF<sup>43</sup>.

**Measurements of the concentration of green-Camuix in cells.** To measure the concentration of mEGFP-CaMKII $\alpha$  or green-Camuix in neurons, we measured the fluorescence intensity of mEGFP at thick dendrites of neurons under two-photon microscopy relative to the brightness of purified mEGFP (1  $\mu$ M) measured under the same imaging setup. When measuring objects larger than two-photon excitation volume ( $\sim 0.9$  femtolitre), the brightness should be proportional to the concentration of fluorophores. We overexpressed mEGFP tagged with poly-histidine tag (His<sub>6</sub>) (pRSET-His<sub>6</sub>-mEGFP) in *Escherichia coli* and purified them using a Nickel-NTI affinity column. The concentration of the purified protein was measured by the absorbance of mEGFP ( $A_{489} = 56,000 \text{ cm}^{-1} \text{ M}^{-1}$ )<sup>44</sup>. For green-Camuix, because FRET reduces its green fluorescence, we corrected for this effect by multiplying  $\tau_{mEGFP}/\tau_{\text{green-Camuix}} \sim 1.4$ , in which  $\tau_{mEGFP}$  and  $\tau_{\text{green-Camuix}}$  are the fluorescence lifetime of mEGFP (2.6 ns) and green-Camuix (1.8 ns), respectively.

**In vitro fluorescence lifetime assay.** To measure the fluorescence lifetime of green-Camuix in cell lysates (Supplementary Fig. 1), we transfected HEK-293 cells with green-Camuix using Lipofectamine 2000. Cells were lysed with a solution containing 40 mM HEPES (pH 8.0), 0.1 mM EGTA, 5 mM magnesium acetate, 0.01% Tween-20, and concentrated with a centrifugal filter (Microcon, Millipore) to remove CaM and ATP. The lysate was diluted 15 times in 136.5 mM potassium gluconate, 17.5 mM KCl, 9 mM NaCl, 1 mM MgCl<sub>2</sub>, 0.2 mM EGTA and 10 mM HEPES-KOH, pH 7.2. We measured the fluorescence lifetime in response to increased free [Ca<sup>2+</sup>] by adding 0.5 mM CaCl<sub>2</sub> and then reduced free [Ca<sup>2+</sup>] by adding 2 mM EGTA.

**Immunofluorescence imaging.** Slices were transfected with green-Camuix (or mEGFP-CaMKII $\alpha$ ) and empty vector, which substitutes mCherry. In parallel experiments shRNA against CaMKII $\alpha$  (35  $\mu$ g) and mEGFP (16  $\mu$ g) were used to verify the specificity of immunostaining. Slices were fixed in 4% paraformaldehyde plus 4% sucrose (in 0.1 M phosphate buffer) for 15 min, followed by incubation in 30% sucrose for 15 min at room temperature. Slices were then re-sectioned using a vibratome at a thickness of 25  $\mu$ m. The sections were permeabilized and blocked overnight at 4 °C in 1% Triton X-100 plus 5% normal goat serum, followed by 30 min blocking in 10% goat serum (in 0.1 M phosphate buffer). Immunostaining was performed for 24 h at 4 °C using as primary antibody mouse anti-CaMKII $\alpha$  (GeneTex Inc.) diluted 1:400 in 5% goat serum in PBS-T (PBS plus 0.1% Triton X-100). After three washes (20 min each) in PBS-T, sections were

incubated with a Texas-red conjugated goat anti-mouse secondary antibody (Invitrogen/Molecular Probes) diluted 1:400 in 10% normal goat serum in PBS-T. The immunofluorescence was quantified under two-photon microscopy in the cell body excluding the nucleus (Supplementary Fig. 2). The value was subtracted by the dark noise of photomultiplier. There was a few per cent of bleedthrough of mEGFP fluorescence to the red channel. To correct this effect, we measured the ratio ( $r$ ) between the red and green fluorescence of mEGFP fluorescence using a purified mEGFP. Then, the bleedthrough was corrected as:  $I_{\text{Texas-red}} = I_{\text{red}} - rI_{\text{green}}$ , in which  $I_{\text{red}}$  and  $I_{\text{Texas-red}}$  are the immunofluorescence before and after the correction, and  $I_{\text{green}}$  is mEGFP fluorescence in the green channel.

**Immunoprecipitation.** HEK-293 cells were transfected with green-Camui $\alpha$  and/or CaMKII $\alpha$  using Lipofectamine 2000. Cells were lysed in a buffer containing 20 mM Tris-HCl, pH 7.4, 137 mM NaCl, 0.5% deoxycholic acid, 25 mM  $\beta$ -glycerophosphate, 1 mM sodium orthovanadate, 2 mM sodium pyrophosphate, 2 mM EDTA, 1 mM sodium fluoride, 0.5 mM dithiothreitol, EDTA-free protease inhibitor cocktail (Roche) and incubated for 30 min at 4 °C. The supernatants were collected after a 30 min microfugal centrifugation at 14,000g at 4 °C. After precleaning with protein A-Sepharose (Sigma), samples were incubated with 5  $\mu$ g per sample of mouse monoclonal anti-Green Fluorescent Protein antibody (MBL) or mouse non-immune IgG (Jackson Laboratories) at 4 °C overnight. The immunocomplexes were precipitated with protein G-Sepharose beads (Sigma) for 1 h at 4 °C and analysed by western blotting (Supplementary Fig. 9).

**Fluctuation correlation spectroscopy.** Fluctuation correlation analysis was performed under a two-photon microscope with a Ti:Sapphire laser tuned at 900 nm at 25–27 °C<sup>45</sup>. Fluorescence signal was detected with a PMT (Hamamatsu), and acquired with a PCI board (Becker-Hickl). HEK-293 cells were transfected by Lipofectamine 2000, allowed to express for 2 days, lysed in a solution containing 40 mM HEPES, pH 8.0, 0.1 mM EGTA, 5 mM magnesium acetate, 0.01% Tween-20, and diluted 10–100 times in a solution containing 10 mM Na-HEPES, pH 7.4, 150 mM NaCl and 0.1 mM EGTA.

41. Nagai, T. *et al.* A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nature Biotechnol.* **20**, 87–90 (2002).
42. Kwok, S. *et al.* Genetically encoded probe for fluorescence lifetime imaging of CaMKII activity. *Biochem. Biophys. Res. Commun.* **369**, 519–525 (2008).
43. Yasuda, R. *et al.* Imaging calcium concentration dynamics in small neuronal compartments. *Sci. STKE* **2004**, pl5 (2004).
44. Tsien, R. Y. The green fluorescent protein. *Annu. Rev. Biochem.* **67**, 509–544 (1998).
45. Schille, P., Haupts, U., Maiti, S. & Webb, W. W. Molecular dynamics in living cells observed by fluorescence correlation spectroscopy with one- and two-photon excitation. *Biophys. J.* **77**, 2251–2265 (1999).

# Design and engineering of an O<sub>2</sub> transport protein

Ronald L. Koder<sup>1\*†</sup>, J. L. Ross Anderson<sup>1\*</sup>, Lee A. Solomon<sup>1</sup>, Konda S. Reddy<sup>1</sup>, Christopher C. Moser<sup>1</sup> & P. Leslie Dutton<sup>1</sup>

**The principles of natural protein engineering are obscured by overlapping functions and complexity accumulated through natural selection and evolution. Completely artificial proteins offer a clean slate on which to define and test these protein engineering principles, while recreating and extending natural functions. Here we introduce this method with the design of an oxygen transport protein, akin to human neuroglobin. Beginning with a simple and unnatural helix-forming sequence with just three different amino acids, we assembled a four-helix bundle, positioned histidines to bis-histidine ligate haems, and exploited helical rotation and glutamate burial on haem binding to introduce distal histidine strain and facilitate O<sub>2</sub> binding. For stable oxygen binding without haem oxidation, water is excluded by simple packing of the protein interior and loops that reduce helical-interface mobility. O<sub>2</sub> affinities and exchange timescales match natural globins with distal histidines, with the remarkable exception that O<sub>2</sub> binds tighter than CO.**

It has long been recognized that natural selection and evolution build complexity into natural proteins and biological systems<sup>1–3</sup>. This complexity frustrates biochemists seeking to understand structure and function<sup>4</sup> and presents an extraordinary challenge to protein engineers who aim to reproduce or create new functions in proteins. So far, this complexity has severely constrained the ability of protein engineers to approach the efficiency of natural protein catalysts<sup>5–10</sup>. However common it may be in nature, we maintain that complexity is not an essential feature of protein as a material, nor is it an essential feature of catalysis, as shown by synthetic chemical systems<sup>11</sup>. By understanding the origins of complexity and making purposeful efforts to separate multiple utilities and minimize complexity during the design and testing of artificial proteins that are completely independent of natural selection, we show how to progressively build in sophisticated biochemical features that reproduce and exceed natural protein function. Our approach follows that long used by artists and architects who develop maquettes—simple models that are progressively altered to test and determine the ultimate characteristics of their constructions.

Protein function requires more than a static structure<sup>12</sup>. In natural proteins, the motion that is part of engineering of protein function is often specific, which can make re-engineering motion for new functions prohibitively difficult. In contrast, artificial proteins offer a full palette of motions that engineering can edit to facilitate those that are productive and remove those that are unproductive.

The engineering-based design of functional synthetic proteins progresses through four stages (Fig. 1): (1) assembly of a simple, robust generic protein framework, such as a helical bundle, of appropriate size to sustain eventual cofactor binding and catalytic function; (2) insertion of cofactor-binding amino acids, keeping the number of amino acid changes low to control complexity; (3) adjusting the sequence for improved structural resolution; and (4) iteratively testing, redesigning and adding engineering elements to refine function.

## Assembly of scaffold and cofactors

Figure 1 describes our starting maquette comprising polar glutamate (E) and lysine (K), as well as non-polar leucine (L). These have high

α-helix-forming propensities<sup>13</sup> which, when arranged in a near-repeating heptad sequence LEELLKK LEELLKL, spontaneously assemble into a water-soluble four-α-helical bundle with glutamate and lysine exposed and leucine buried in a molten globular interior<sup>14</sup>. To make a bundle of a length typical of natural proteins, we used almost four heptad near-repeats (1 in Fig. 1). Such a bundle is free to associate with helices oriented parallel or anti-parallel. Adding an amino-terminal CGGG sequence for disulphide-mediated dimerization restricts helical topologies to *syn* or *anti*. Completing this stage, we replaced an internal leucine at e-position 7 of each helix with tryptophan to facilitate optical detection of the protein.

In the second stage of design, we found that replacing leucines at internal a-positions 10 and 24 with histidine was sufficient to anchor up to four haems in the bundle<sup>15</sup>. We also replaced a-position leucine 17 between the haems with phenylalanine, an amino acid commonly found near haems in natural proteins. To allow us to discriminate between the haem-binding sites, we replaced another interior leucine with arginine (2 in Fig. 1). No effort was made to design core cavities that accommodate haem; reliance was put on interior histidine—iron ligation to position the haem in a malleable hydrophobic interior.

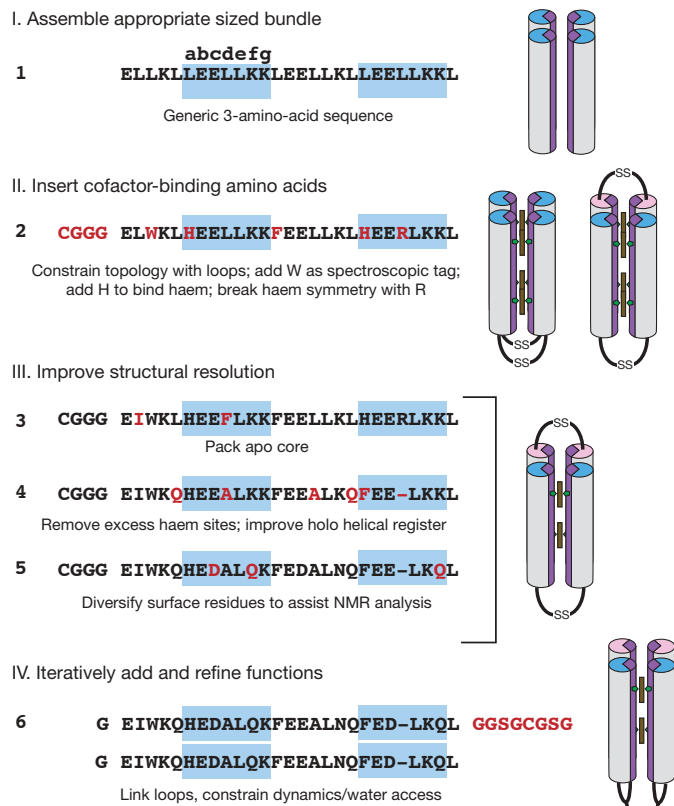
Structural resolution is helpful in keeping progressive design on track. Singular structures in these bundles can be engineered without computation; introduction of β-branched aliphatic or aromatic residues along with a polar bond across helices provided by the histidines (3 in Fig. 1) confers tertiary structure to the four-α-helix interior of the apo protein as seen by NMR and X-ray crystallography<sup>16,17</sup>.

Although various functions can be added to this basic haem-binding protein, for oxygen transport we elected to simplify the design by lowering the haem capacity from four to two by replacing histidine at position 24 with phenylalanine. Inspection of model apo structures showed that a substantial rotation of >50° around the helical axes was required on haem binding to accommodate the histidine rotamers typical of natural bis-histidine haem-binding proteins (Supplementary Fig. 1)<sup>18</sup>. This rotation exports hydrophobic interior residues into, and imports polar residues from, the aqueous phase. Modelling identified four amino acids for substitution to interfacially compatible

<sup>†</sup>The Johnson Research Foundation, Department of Biochemistry and Biophysics, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. <sup>†</sup>Present address: Department of Physics, The City College of New York, New York 10031, USA.

\*These authors contributed equally to this work.





**Figure 1 | The design of an artificial oxygen transport protein (6).** Design begins with an extremely simple heptad near-repeat sequence using just three amino acids (1). Blue highlight marks heptad divisions. Design progresses through stages of amino acid changes (red) and trial intermediates (2–5) that are tested to reveal functional properties and to clarify the roles of individual amino acids. In 2–5, helical sequences shown are linked by cysteine disulphide loops and self-assemble in four-helix bundles, whereas in 6 a longer loop (red) unites two identical sequences, with the loops themselves now disulphide-linked, as shown to the right.

alanines or glutamines<sup>19</sup> and one for deletion to more favourably realign the binary pattern after haem addition, yielding 4 (Fig. 1). However, three inwardly rotating b-position glutamates at positions 11, 18 and 25 were deliberately left in place<sup>18</sup> to apply strain to weaken one of the two histidine–haem iron ligation interactions, as occurs in neuroglobin. This creates an entatic state<sup>20</sup>. Fourier transform infrared spectroscopy (FTIR) shows that these strain-inducing glutamates have pK values (that is, pH values at which the residues are half protonated) increased by more than two units (Supplementary Fig. 2) and, as in natural haem proteins, are strongly electrostatically coupled to haem oxidation–reduction, changing the pK by more than three units<sup>21</sup>. NMR showed that addition of two haem Bs transformed unstructured apo-4 into a well-defined tertiary structure<sup>22</sup> (Supplementary Fig. 3). It appears that the histidine–iron polar bond can also provide a nucleus for interior packing that promotes a singular structure around different porphyrin cofactors<sup>22</sup>. To ease NMR structure determination and assignment of 90% of the peptide backbone (Supplementary Fig. 4), external residues were also diversified at this time, yielding 5 (Fig. 1).

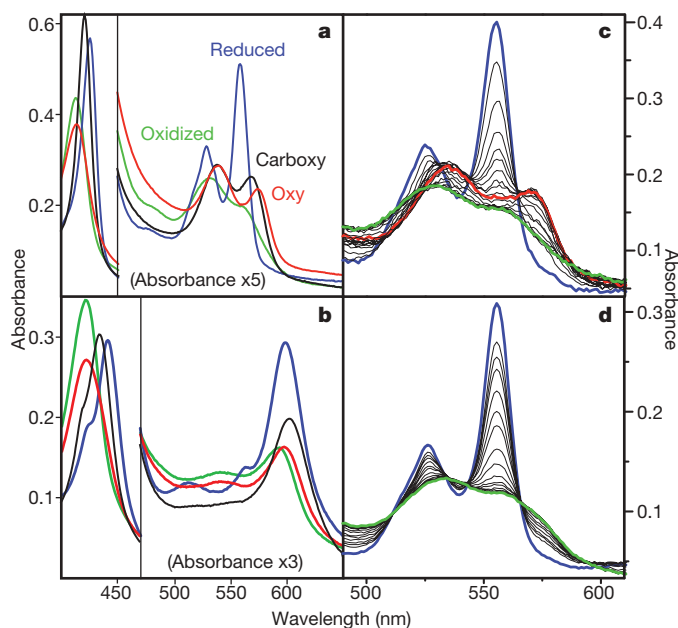
### Exclusion of water for oxy-ferrous haem stability

None of the molecules discussed to this point, indeed no artificial haem proteins so far examined<sup>23–25</sup>, bind O<sub>2</sub> stably at room temperature (23 °C). This is not because the oxy-ferrous haem cannot be formed. Structures 2, 3, 4 and 5 all include the glutamate-based helical strain, which weakens the histidine ligation and allows competition from other ligands. By first exposing these haem proteins to CO and then illuminating at a low temperature to initiate ligand exchange for O<sub>2</sub>, as was done previously with cytochrome oxidase<sup>26</sup>, each forms an

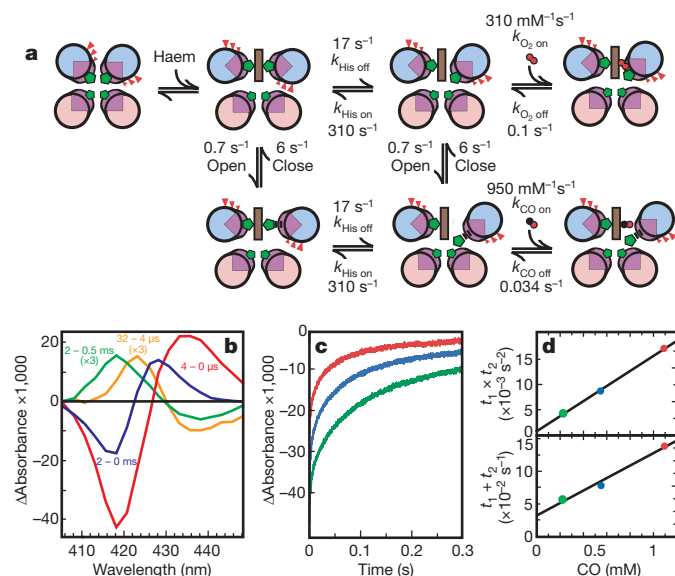
oxy-ferrous haem at near 100% yield at –15 °C. The probable reason for room temperature instability is the access of water and protons to the haem-binding site, which facilitates electron transfer from ferrous haem to oxygen. This is clear from mutant studies of globins and the relative stability of the oxy-ferrous haem in chemical systems in water-free solvents<sup>11,27,28</sup>.

To engineer water exclusion we constrained the culpable protein motion. The crystal and NMR structures of 3 reveal that one of the two interhelical interfaces, that which lies between the helices not constrained by haem binding, has an unusually low degree of surface complementarity and a high degree of inter-helix motion. Indeed, sequence 2 can be modified to introduce large-scale *syn–anti* flipping<sup>29</sup>. To constrain motion, the loops were reconfigured to link the helices across the most mobile interface. This also allowed the loops to be linked into a monomeric ‘candelabra’ structure, further constraining motion (6 in Fig. 1). NMR confirmed the water-restricting effects of loop reconfiguration and monomerization. Hydrogen isotope (H/D) analysis of hydrogen exchange protection factors of 4 and 6 at pH 7 and 25 °C (Supplementary Fig. 5) shows that water-mediated proton exchange is complete within 15 min for 4 and is much slower for 6 with several core residue backbone amides, including some close to the haem, exchanging on a timescale of several hours.

Maquettes 2–6 display ferric and ferrous haem visible spectra indicative of six-coordinate bis-histidine ligated haem B, characteristic of cytochrome *b*, deoxy-neuroglobin and cytoglobin, and quite distinct from the five-coordinate myoglobin and haemoglobin. Because the dissociation constant ( $K_d$ ) for binding of the first haem (<1 nM) is much tighter than that for the second haem (50 nM)<sup>22</sup>, we simplified spectral analysis by binding one haem per bundle (Fig. 2a). NMR assignments unambiguously identified the first haem B to bind at H7 positions at the open end of the candelabra structure of 6. Only ferrous haem 6 shows rapid and complete conversion of the ferrous haem into the oxy-ferrous haem with a half time of ~50 ms measured by stopped-flow spectroscopy. This oxy-ferrous spectrum is remarkably similar to that of native neuroglobin (Supplementary Fig. 6). The oxy-ferrous



**Figure 2 | Haem maquette spectra.** a, b, At –15 °C, these spectra are stable for more than an hour: oxidized (green), reduced (blue), carboxy-ferrous (black) or oxy-ferrous (red) artificial oxygen transport protein 6 with either haem B (a) or haem A (b) as the cofactor. c, d, Stopped-flow spectral changes for mixing the reduced haem B proteins with oxygen at 15 °C. The fully designed oxygen transport protein 6 (c) shows the transformation of the reduced haem (blue) to the oxy-ferrous state (red), which eventually becomes oxidized (green), whereas the early intermediate 2 (d) proceeds directly and rapidly to the oxidized form.



**Figure 3 | Modelling kinetics of haem ligand binding and release.** **a**, Haem binding rotates helices but incurs strain by burying glutamates (red). Blue and pink indicate, respectively, the N- and C-terminal ends of the helices. Like some natural haemoglobins, at least two conformations are present; the closed form cannot bind CO. Histidine (His) release with helical rotation precedes gaseous ligand binding. **b**, Immediate CO photolysis difference spectra (4  $\mu$ s minus 0  $\mu$ s, red) of **6** haem B is followed by a rapid microsecond relaxation (yellow) followed by slower histidine and CO binding (green) and even slower displacement of histidine by CO (blue). Subsequent biphasic recombination kinetics (characteristic times  $t_1$  and  $t_2$ ) at 418 nm (**c**) and the product (top) and sum (bottom) of these characteristic times as a function of CO concentration (**d**) determine histidine on/off rates.

state is stable for tens of seconds before single electron transfer from ferrous haem to  $O_2$  appears to generate superoxide.

A similar experiment with CO shows rapid displacement of one histidine to generate an indefinitely stable carboxy-ferrous haem species in a 400-ms half time (not shown). After CO photolysis, CO rebinding to **6** is multiphasic (Fig. 3 and Supplementary Fig. 7), as observed in the hexacoordinate globins; detailed analysis of these data allows the estimation of the histidine on- and off-rates (Table 1). CO rapid mixing shows slower binding than expected from photolysis, indicating the presence of at least two conformations that differ in CO accessibility, termed 'open' and 'closed' in analysis of hexacoordinate haemoglobins<sup>30</sup>. Saturation of the observed CO binding rate allows estimates of the opening and closing rates (Supplementary Fig. 8). Competition between CO recombination and oxidation by added ferricyanide also allows an estimation of the rate of thermal CO dissociation. Table 1 presents the kinetic constants (also shown in Fig. 3a) and equilibrium dissociation constants that have been determined for **6** and contrasts them with several natural oxygen-binding proteins. Whereas the oxygen off-rate is similar to that of human

neuroglobin<sup>31</sup>, the on-rate is almost 100 times slower and resembles that of *Ascaris* haemoglobin. These artificial proteins are robust and adaptable enough to redesign for the testing of the hypothesis that a large hydrophobic pocket, as in neuroglobin<sup>32</sup>, can speed  $O_2$  binding, whereas proximal strain, as in *Ascaris* haemoglobin<sup>33</sup>, slows  $O_2$  binding.

### Testing engineering elements

To test whether the helical rotation model of histidine strain was indeed operating to promote CO and  $O_2$  binding, we changed the b-position glutamates of **6** to alanines. As anticipated, CO binding slowed by more than an order of magnitude (Supplementary Fig. 9). Moreover,  $O_2$  failed to form detectable oxy-ferrous haem. Instead, the haem underwent oxidation within the mixing time, plausibly by means of an outer-sphere electron transfer. Thus, removal of the modelled entatic state<sup>20</sup> by conversion of interfacial glutamates to alanines disables the histidine for  $O_2$  or CO exchange-gated rotational mechanism.

In exploring the effectiveness of loops to control mobility of free helices and thereby exclude water from the interior, we found that the oxy-ferrous haem is not stabilized by looping together helices already linked by haem (**5**), but is stabilized by looping together helices not linked by haem (**6**). Disulphide linkage of the loops in **6** proves unessential, because eliminating the disulphide through a Cys-to-Ser substitution maintains oxy-ferrous haem stability.

Our development of helical strain and motion constraint to promote  $O_2$  binding to ferrous haem is not confined to one site, but extends along the four-helix bundle to haems at other positions. A di-haem variant of **6** with histidine at position 7 and the complementary position 42, or a single-haem variant with histidine only at position 42, each display oxy-ferrous haem properties similar to that observed in **6** (Supplementary Fig. 10).

### Haem and substrate specificity

The interior of **6** adopts a unique structure not only around haem B but also around other porphyrins including haem A (Supplementary Figs 11 and 12). Haem A has markedly different peripheral substitutions (Fig. 2b), which result in different redox and spectral properties. The oxidation–reduction midpoint potential at pH 8 ( $E_{m8}$ ) measured versus the normal hydrogen electrode (NHE) of bound haem A is  $-100$  mV (data not shown). This is about 200 mV more positive than haem B and hence less favoured to reduce  $O_2$  to  $O_2^{\cdot-}$ . Accordingly, the oxy-ferrous haem A (spectrally analogous to the cytochrome oxidase Compound A described previously<sup>26</sup>) persists longer, about half a minute, before electron transfer.

In all natural haemoglobins with distal histidines, either preferentially bound to the haem iron, as in neuroglobin, or displaced from the iron, as in myoglobin or human haemoglobin, CO is a poison that binds more tightly than  $O_2$ . Our work shows this is not an essential property. The binding of CO to **6** is weaker than to natural globins (Table 1). This results in a net tenfold discrimination favouring  $O_2$  binding over CO binding, the largest observed for any distal histidine haem–protein complex, and comparable to distal tyrosine sites with extreme  $O_2$  affinity. It seems probable that the strained, distal

**Table 1 | Haem iron ligand on- and off-rates and equilibrium constants in natural and artificial proteins**

| Haem protein                             | Ligation              | $k_{\text{His on}} (\text{s}^{-1})$ | $k_{\text{His off}} (\text{s}^{-1})$ | $k_{\text{CO on}} (\mu\text{M}^{-1} \text{s}^{-1})$ | $k_{\text{CO off}} (\text{s}^{-1})$ | $K_{\text{dCO}} (\text{nM})$ | $k_{\text{O}_2 \text{ on}} (\mu\text{M}^{-1} \text{s}^{-1})$ | $k_{\text{O}_2 \text{ off}} (\text{s}^{-1})$ | $K_{\text{dO}_2} (\text{nM})$ | $K_{\text{dO}_2}/K_{\text{dCO}}$ |
|--|-----------------------|-------------------------------------|--------------------------------------|---|-------------------------------------|------------------------------|--|--|-------------------------------|----------------------------------|
| <b>6</b> (haem B)                        | bis-His               | 310                                 | 17                                   | 0.95 (0.006)  | 0.034                               | 36 (6,000)                   | 0.31 (0.017)   | 0.1  | 32 (600)                      | (0.1)                            |
| Neuroglobin <sup>31</sup>                | bis-His               | 2,000                               | 4.5                                  | 65 (0.15)   | 0.014                               | 0.21 (93)                    | 250 (0.56)   | 0.8  | 3.2 (1,400)                   | (15)                             |
| Myoglobin <sup>47</sup>                  | His-aquo (distal His) | n/a                                 | n/a                                  | 0.5   | 0.019                               | 37                           | 14   | 12   | 860                           | 23                               |
| Human haemoglobin <sup>48</sup>          | His-aquo (distal His) | n/a                                 | n/a                                  | 2.2   | 0.009                               | 4                            | 19   | 15   | 770                           | 190                              |
| <i>Ascaris</i> haemoglobin <sup>49</sup> | His-aquo (distal Tyr) | n/a                                 | n/a                                  | 0.2   | 0.018                               | 90                           | 1.5  | 0.004  | 3                             | 0.03                             |
| Microperoxidase <sup>50</sup>            | His-aquo              | n/a                                 | n/a                                  | 20  | 0.01                                | 0.5                          |  |  |                               |                                  |
| Combinatorial bundles <sup>44</sup>      | Mixed                 | ?                                   | ?                                    | (3–11)  | 0.03–0.11                           | (6–25)                       |  |  |                               |                                  |

Apparent binding rates (in parentheses) are slowed by bis-histidine (bis-His) ligation. Corresponding rates not in parentheses are estimated binding rates without histidine impedance facilitating comparison with proteins with different fractions of bis-histidine ligation. His-aquo has water as the sixth iron ligand. n/a, not applicable; ?, unknown.

histidine in **6** remains available to stabilize bound oxygen, as is seen in haemoglobins, myoglobins and horseradish peroxidase<sup>34</sup>.

### Protein engineering to control complexity and utility

Management of complexity and multiple-utility is critical to engineering protein design. Natural proteins accumulate complexity in a manner analogous to that described in genetic systems as 'Muller's ratchet': a change is made, typically for minor or even no selective advantage, that becomes essential as new changes come to depend on this old change. The resulting interdependency between amino acids is the source of protein complexity. Not only does one amino acid commonly affect the function of nearby or allosterically remote amino acids, any amino acid typically supports multiple functions or utilities, in a way analogous to the organism scale description of Darwin<sup>1,35,36</sup>.

Amino acid interdependency and multiple-utility of amino acids presents a serious challenge to re-engineering a natural protein for a new purpose. When more than a few amino acids are 'rationally' introduced to natural proteins by site-directed mutagenesis, protein stability and activity are often deeply compromised<sup>37</sup>. Limited adaptability of natural proteins to accommodate new catalytic site functionalities has been the experience so far with a variety of powerful approaches including abzymes<sup>38</sup>, computational design combined with directed evolution<sup>6</sup> and recombination through domain swapping<sup>39</sup>.

Maquette design that begins with simple protein sequences never exposed to natural selection and tests the effects of amino acid changes through the progressive stages of design does not eliminate complexity or multiple utility, but allows them to be more clearly seen and moderated. This approach offers reconciliation of each amino acid with specific roles in structure and function, the choice to eliminate redundancies, and the freedom to organize assembly and mechanistic elements while incorporating the next stage of development. As a result, it takes relatively few amino acid changes to incorporate the function of oxygen transport into a generic three-amino-acid bundle. The fact that such a modest redesign succeeds supports the view that many protein-engineering elements do not require atomistic precision and that exacting mimics of natural protein sites is neither necessary nor good engineering. Indeed, despite the widespread fragility of enzymes to multiple active site mutations, it appears that relatively simple and unsophisticated engineering principles underlie natural protein functions such as electron transfer<sup>40</sup> and even catalysis<sup>41,42</sup>.

One of these key engineering principles is the creation and control of protein motion. The long-range helical motion and strain that facilitates ligand exchange for oxygen transport, while still permitting relatively secure haem binding by means of bis-histidine ligation, would not have been possible merely by targeting the static structure of the final oxy-ferrous state in a natural haem protein. Any static structure-focused design<sup>6,7,38</sup> fails to address the molecular choreography often critical for function. In contrast, liberal motion is equally unworkable. Our balance of allowing inter-helical protein motion but constraining it through bonded and loop contacts allows small ligands to penetrate and bind while excluding the water that so profoundly affects active site properties. In different maquette variants, limiting water access slowed the rate for haem-oxygen electron transfer by four orders of magnitude, suggesting the raising of a catalytic barrier by >5 kcal.

By keeping the maquette design simple and complexity low, we maintain the capacity to intentionally insert more complex functional elements such as allostery and cooperativity. The helical rotation built into this system introduces a long-range negative cooperativity in the two haem-binding sites. Versions of **6** that modify the histidines at one or the other binding site clearly show that on their own both haem-binding sites have high affinity, whereas, together, rotation on binding the first haem weakens the second. Such long-range helical coupling is a prospective means to build a maquette with positive cooperativity between the haem sites for oxygen binding, recreating the vital oxygen transport regulation for which haemoglobin is so well known.

The creation of the first completely artificial oxygen transport protein allows us to reconsider the design of natural oxygen transport globins. Despite the common view that natural globins are full of exquisitely refined functional properties reflected in the globin fold<sup>43</sup>, it is clear that transport function is readily achieved without such a fold. Indeed, the results presented here reinforce the developing view that the physical chemistry of haem in an oxygen transport protein can accommodate a wide range of histidine, O<sub>2</sub> and CO ligand exchange rates including, as we have seen here, the preference for O<sub>2</sub> over CO binding. The discrimination favouring O<sub>2</sub> over CO binding comes about without added engineering. Taking into account the impedance from the 95% bis-histidine ligation in **6** expected from the histidine on- and off-rate ratio, binding of CO and O<sub>2</sub> ligands are both relatively slow compared to most natural globins, but the thermal release rates are towards the fast and slow ends of the natural range, respectively (Table 1)<sup>44</sup>. Traditionally, slow O<sub>2</sub> off-rates are explained in terms of nearby polar residues providing a hydrogen bond to the O<sub>2</sub>; simple modelling of **6** shows the distal histidine could play this part here. However, no special design was required beyond giving this histidine sufficient mobility by means of the helical strain arising from glutamate burial. The ease with which globin-like properties can be reproduced in a completely unrelated and simply engineered maquette indicates that the relatively complex globin fold is for the most part unremarkable, and may be common in nature not because of a uniquely capable design for oxygen binding, but simply because it is good enough.

### METHODS SUMMARY

Protein expression, purification and confirmation of purity were performed as described previously<sup>22</sup>. CO rebinding after photolysis by a frequency-doubled YAG laser at 532 nm, ~1-ns pulse width (Fig. 3c), is multi-exponential, a characteristic signature of human neuroglobins<sup>45</sup>. Detailed analysis of time-resolved spectra in the flash experiments is consistent with a rapid exponential phase due to a change in pentacoordinate haem conformation (see Fig. 3b), similar to that reported for myoglobin<sup>46</sup>. After this relaxation, slower biphasic kinetics are associated with direct CO and histidine binding, followed by final histidine-CO exchange. Using the kinetic model of Hargrove<sup>45</sup> to analyse the CO-concentration-dependence of the CO binding rates on flash photolysis allows the estimation of histidine on- and off rates and per cent bis-histidine ligation for ferrous **6** (Fig. 3d).

Predominantly mono-exponential binding kinetics of O<sub>2</sub> and CO with the ferrous haem B in **6** were followed spectroscopically in rapid stopped-flow mixing over gas concentrations from 2% to 50% saturation at 15 °C using an Olis RSM 1000 spectrometer that scans both the alpha band and Soret band absorptions 1,000 times a second followed by SVD analysis of the kinetic components. The dissociation rate of the CO complex was determined by competition with ferricyanide using a range of ferricyanide concentrations according to the ferric haem trapping method described previously<sup>44</sup>.

Preparations of ferrous haem synthetic proteins in CO saturated 30% ethylene glycol borate buffer at pH 9.0 are stable to ligand exchange on a timescale of several minutes in the dark at -15 °C. This allows spectra such as those shown in Fig. 2 to be prepared by re-equilibration with O<sub>2</sub> or O<sub>2</sub>/CO gas mixtures followed by photolysis by halogen lamp illumination for several seconds to initiate rapid O<sub>2</sub> binding, as described previously<sup>26</sup>.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 28 October 2008; accepted 27 January 2009.

1. Darwin, C. *Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life* 6th edn (Murray, 1872).
2. Muller, H. J. The relation of recombination to mutational advance. *Mutat. Res.* **1**, 2-9 (1964).
3. Csete, M. E. & Doyle, J. C. Reverse engineering of biological complexity. *Science* **295**, 1664-1669 (2002).
4. Kraut, D. A., Carroll, K. S. & Herschlag, D. Challenges in enzyme mechanism and energetics. *Annu. Rev. Biochem.* **72**, 517-571 (2003).
5. Bolon, D. N. & Mayo, S. L. Enzyme-like proteins by computational design. *Proc. Natl Acad. Sci. USA* **98**, 14274-14279 (2001).
6. Jiang, L. et al. *De novo* computational design of retro-aldol enzymes. *Science* **319**, 1387-1391 (2008).
7. Rothlisberger, D. et al. Kemp elimination catalysts by computational enzyme design. *Nature* **453**, 190-195 (2008).



8. Kaplan, J. & DeGrado, W. F. *De novo* design of catalytic proteins. *Proc. Natl Acad. Sci. USA* **101**, 11566–11570 (2004).
9. Moffet, D. A. *et al.* Peroxidase activity in heme proteins derived from a designed combinatorial library. *J. Am. Chem. Soc.* **122**, 7612–7613 (2000).
10. Monien, B. H. *et al.* Detection of heme oxygenase activity in a library of four-helix bundle proteins: towards the *de novo* synthesis of functional heme proteins. *J. Mol. Biol.* **371**, 739–753 (2007).
11. Collman, J. P., Boulatov, R., Sunderland, C. J. & Fu, L. Functional analogues of cytochrome c oxidase, myoglobin, and hemoglobin. *Chem. Rev.* **104**, 561–588 (2004).
12. Jencks, W. P. Binding-energy, specificity, and enzymic catalysis – circe effect. *Adv. Enzymol.* **43**, 219–410 (1975).
13. Chou, P. Y. & Fasman, G. D. Empirical predictions of protein conformation. *Annu. Rev. Biochem.* **47**, 251–276 (1978).
14. Regan, L. & DeGrado, W. F. Characterization of a helical protein designed from 1st principles. *Science* **241**, 976–978 (1988).
15. Robertson, D. E. *et al.* Design and synthesis of multi-heme proteins. *Nature* **368**, 425–431 (1994).
16. Gibney, B. R. *et al.* Iterative protein redesign. *J. Am. Chem. Soc.* **121**, 4952–4960 (1999).
17. Huang, S. S. *et al.* X-ray structure of a maquette scaffold. *J. Mol. Biol.* **326**, 1219–1225 (2003).
18. Huang, S. S. *et al.* The HP-1 maquette: from an apoprotein structure to a structured hemoprotein designed to promote redox-coupled proton exchange. *Proc. Natl Acad. Sci. USA* **101**, 5536–5541 (2004).
19. Marshall, S. A. & Mayo, S. L. Achieving stability and conformational specificity in designed proteins via binary patterning. *J. Mol. Biol.* **305**, 619–631 (2001).
20. Vallee, B. L. & Williams, R. J. P. Metalloenzymes – entatic nature of their active sites. *Proc. Natl Acad. Sci. USA* **59**, 498–505 (1968).
21. Shifman, J. M. *et al.* Functionalized *de novo* designed proteins: mechanism of proton coupling to oxidation/reduction in heme protein maquettes. *Biochemistry* **37**, 16815–16827 (1998).
22. Koder, R. L. *et al.* Native-like structure in designed four helix bundles driven by buried polar interactions. *J. Am. Chem. Soc.* **128**, 14450–14451 (2006).
23. Isogai, Y. *et al.* Design and synthesis of a globin fold. *Biochemistry* **38**, 7431–7443 (1999).
24. Gibney, B. R. *et al.* Self-assembly of heme A and heme B in a designed four-helix bundle: implications for a cytochrome c oxidase maquette. *Biochemistry* **39**, 11041–11049 (2000).
25. Zhuang, J. Y. *et al.* Design of a five-coordinate heme protein maquette: a spectroscopic model of deoxymyoglobin. *Inorg. Chem.* **43**, 8218–8220 (2004).
26. Chance, B., Saronio, C. & Leigh, J. S. Functional intermediates in reaction of cytochrome-oxidase with oxygen. *Proc. Natl Acad. Sci. USA* **72**, 1635–1640 (1975).
27. Shikama, K. The molecular mechanism of autoxidation for myoglobin and hemoglobin: a venerable puzzle. *Chem. Rev.* **98**, 1357–1373 (1998).
28. Wang, J. H. Hemoglobin studies. 2. A synthetic material with hemoglobin-like property. *J. Am. Chem. Soc.* **80**, 3168–3169 (1958).
29. Grosset, A. M. *et al.* Proof of principle in a *de novo* designed protein maquette: an allosterically regulated, charge-activated conformational switch in a tetra- $\alpha$ -helix bundle. *Biochemistry* **40**, 5474–5487 (2001).
30. Trent, J. T., Hvitved, A. N. & Hargrove, M. S. A model for ligand binding to hexacoordinate hemoglobins. *Biochemistry* **40**, 6155–6163 (2001).
31. Dewilde, S. *et al.* Biochemical characterization and ligand binding properties of neuroglobin, a novel member of the globin family. *J. Biol. Chem.* **276**, 38949–38955 (2001).
32. Pesce, A. *et al.* Human brain neuroglobin structure reveals a distinct mode of controlling oxygen affinity. *Structure* **11**, 1087–1095 (2003).
33. Peterson, E. S. *et al.* A comparison of functional and structural consequences of the tyrosine B10 and glutamine E7 motifs in two invertebrate hemoglobins (*Ascaris suum* and *Lucina pectinata*). *Biochemistry* **36**, 13110–13121 (1997).
34. Borovik, A. S. Bioinspired hydrogen bond motifs in ligand design: the role of noncovalent interactions in metal ion mediated activation of dioxygen. *Acc. Chem. Res.* **38**, 54–61 (2005).
35. McLendon, G. Control of biological electron-transport via molecular recognition and binding – the velcro model. *Struct. Bond.* **75**, 159–174 (1991).
36. Page, C. C., Moser, C. C. & Dutton, P. L. Mechanism for electron transfer within and between proteins. *Curr. Opin. Chem. Biol.* **7**, 551–556 (2003).
37. Brannigan, J. A. & Wilkinson, A. J. Protein engineering 20 years on. *Nature Rev. Mol. Cell Biol.* **3**, 964–970 (2002).
38. Hilvert, D. Critical analysis of antibody catalysis. *Annu. Rev. Biochem.* **69**, 751–793 (2000).
39. Carbone, M. N. & Arnold, F. H. Engineering by homologous recombination: exploring sequence and function within a conserved fold. *Curr. Opin. Struct. Biol.* **17**, 454–459 (2007).
40. Moser, C. C. *et al.* Nature of biological electron-transfer. *Nature* **355**, 796–802 (1992).
41. Benkovic, S. J. & Hammes-Schiffer, S. A perspective on enzyme catalysis. *Science* **301**, 1196–1202 (2003).
42. Warshel, A. Computer simulations of enzyme catalysis: methods, progress, and insights. *Annu. Rev. Biophys. Biomol. Struct.* **32**, 425–443 (2003).
43. Frauenfelder, H., McMahon, B. H. & Fenimore, P. W. Myoglobin: the hydrogen atom of biology and a paradigm of complexity. *Proc. Natl Acad. Sci. USA* **100**, 8615–8617 (2003).
44. Moffet, D. A. *et al.* Carbon monoxide binding by *de novo* heme proteins derived from designed combinatorial libraries. *J. Am. Chem. Soc.* **123**, 2109–2115 (2001).
45. Hargrove, M. S. A flash photolysis method to characterize hexacoordinate hemoglobin kinetics. *Biophys. J.* **79**, 2733–2738 (2000).
46. Ansari, A. *et al.* The role of solvent viscosity in the dynamics of protein conformational-changes. *Science* **256**, 1796–1798 (1992).
47. Springer, B. A., Sligar, S. G., Olson, J. S. & Phillips, G. N. Mechanisms of ligand recognition in myoglobin. *Chem. Rev.* **94**, 699–714 (1994).
48. Mathews, A. J. *et al.* The effects of E7 and E11 mutations on the kinetics of ligand-binding to R-state human-hemoglobin. *J. Biol. Chem.* **264**, 16573–16583 (1989).
49. Goldberg, D. E. Oxygen-avid hemoglobin of *Ascaris*. *Chem. Rev.* **99**, 3371–3378 (1999).
50. Sharma, V. S., Schmidt, M. R. & Ranney, H. M. Dissociation of CO from carboxyhemoglobin. *J. Biol. Chem.* **251**, 4267–4272 (1976).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank A. J. Wand for assistance in NMR measurements, P. R. Rich for FTIR measurements, M. S. Hargrove for neuroglobin reference spectra and discussions, and D. Hilvert for suggestions. This work was supported by grants from US Department of Energy, US National Institute of Health and US National Science Foundation as detailed in Supplementary Information.

**Author Contributions** R.L.K. and J.L.R.A. both designed proteins and performed the bulk of the measurements; K.S.R. made initial spectroscopic observations and L.A.S. contributed to spectroscopic measurements; and C.C.M. designed and performed spectroscopic measurements and analysis. Paper preparation was largely conducted by C.C.M. and P.L.D.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to P.L.D. ([dutton@mail.med.upenn.edu](mailto:dutton@mail.med.upenn.edu)).

## METHODS

Typical CO photolysis conditions were 10  $\mu$ M synthetic protein in 20 mM borate, pH 9.0, 100 mM KCl and CO/Ar gas mixtures ranging from 4% to 100% CO gas saturation with a slight excess of sodium dithionite monitored in the ultraviolet to ensure that the haem was reduced.

Table 1 reports the direct, apparent CO and O<sub>2</sub> binding rates as well as estimates of the faster rates when not bis-histidine ligated. Apparent CO binding rates from photolysis are ~20 times faster than that reported by stopped flow, indicating that shortly after photolysis bis-histidine ferrous **6** may be in a relatively unrelaxed 'open' state, prone to histidine dissociation and CO rebinding, compared to the dark, equilibrium, resting 'closed' state<sup>30</sup>.

Stable spectra of the oxy-ferrous form are also obtained by direct bubbling of ferrous haem protein with oxygen at -15 °C without using CO as an intermediate. Competition between CO and O<sub>2</sub> in a range of mixtures of the two gases reaches equilibrium over 30 min after initial photolysis at -15 °C and indicates that O<sub>2</sub> binds with 10 times greater affinity than CO (for example, a 91:9 ratio of CO:O<sub>2</sub> gives 50:50 mix of bound CO and O<sub>2</sub>).

**Materials.** Peptide synthesis reagents were purchased from PerSeptive Biosystems. Haemin was from Fluka. The chloride salt of Fe<sup>3+</sup>-protoporphyrin III was synthesized by methods described previously<sup>51</sup>. Deuterium oxide, <sup>15</sup>N-ammonium chloride, <sup>13</sup>C-glucose and (trimethylsilyl)-propionate were from Cambridge Isotope Laboratories. All other solvents and reagents were from either Fisher Scientific or Sigma.

**Peptide synthesis and purification.** Solid-phase syntheses of **1–5** were performed on a continuous flow PerSeptive Biosystems as described previously. Crude peptides were purified to homogeneity by reversed phase C<sub>18</sub> HPLC using aqueous-acetonitrile gradients containing 0.1% (vol/vol) trifluoroacetic acid. Peptide homogeneity and composition were assayed by analytical HPLC and laser desorption mass spectrometry.

The gene encoding **6** was designed using the program Amplify<sup>52</sup>. The gene was constructed using assembly PCR<sup>53</sup> and ligated into the overexpression vector pET 32d(+) (Novagen, Inc.) modified to replace the enterokinase cleavage site with a TEV-protease cleavage site, yielding the plasmid pH7-TEV. The plasmid was transformed into DH5 $\alpha$  cells and sequenced in the region of interest to verify the presence of the intact gene, and the plasmid was transformed into BL21(DE3) cells. Mutations to **6** were created using the Quick Change protocol (Stratagene, Inc.).

For unlabelled expression, these were grown in TPP medium<sup>54</sup> at 37 °C to a *D*<sub>600</sub> of 1.0 and induced with IPTG at a final concentration of 1 mM for 2 h before collection. For <sup>15</sup>N-labelled expression, cells were grown at 37 °C in minimal

media containing 1 g l<sup>-1</sup> <sup>15</sup>N ammonium chloride to a *D*<sub>600</sub> of ~1.0, induced with 1 mM IPTG, and shaken at 37 °C for an additional 3 h. For <sup>13</sup>C,<sup>15</sup>N-double labelling, cells were grown in minimal media containing 1% w/v <sup>13</sup>C glucose, 1 g l<sup>-1</sup> <sup>15</sup>N ammonium chloride and 8 ml l<sup>-1</sup> <sup>13</sup>C,<sup>15</sup>N Bioexpress medium (Cambridge Isotope Labs). Cells were grown at 37 °C to a *D*<sub>600</sub> of ~1.0, induced with 1 mM IPTG, and shaken at 37 °C for an additional 2 h.

Cells were collected by centrifugation, broken open using a French press, and purified on a Ni-nitrilotriacetic acid column (Qiagen, Inc.) according to the manufacturer's instructions. The fusion protein was dialysed into 50 mM Tris-HCl, 1 mM DTT, pH 8.0, and then cleaved overnight (15–18 h) with His<sub>6</sub>-tagged TEV protease (Invitrogen, Inc.). The reaction mixture was filtered through Ni-nitrilotriacetic acid resin, and the purified **6** concentrated by lyophilization. Purified peptides were dissolved in 20 mM K<sub>2</sub>HPO<sub>4</sub> and 100 mM KCl, pH 8.0, and cysteine residues were air-oxidized to the symmetric disulphides overnight. Disulphide formation was followed by analytical C<sub>18</sub> HPLC.

**FTIR.** **6** was washed by dilution with 100 mM potassium phosphate and 100 mM KCl at pH 6.0 or pH 8.0 and concentrated to ~1 mM with Vivascience centrifugal concentrators. Benzyl viologen (100  $\mu$ M) was added as a redox mediator and the sample was placed in a microelectrochemical chamber above an ATR-FTIR microprism (3 bounce silicon; 3 mm diameter; SensIR). The chamber had a 5-mm-diameter glassy carbon working electrode within 1 mm of the prism surface and the sample solution was connected to a counter electrode and an Ag/AgCl reference electrode via an ion-permeable frit. Potential was switched cyclically between 0 mV and -450 mV (versus SHE). Before each potential change, a reference scan was recorded. The potential was then switched and a sample spectrum was recorded after 10 min to allow for redox equilibration. Data at pH 6.0 average 50 reductive minus 50 oxidative cycles of 500 interferograms each at 4 cm<sup>-1</sup> resolution; data at pH 8 average 40 reduced minus oxidized cycles; the double difference spectrum pH 6 minus 0.58 pH 8 reflected appropriate protein concentration-dependent weighting at the two pH values.

51. Smith, K. M., Parish, D. W. & Inouye, W. S. Methyl deuteration reactions in vinylporphyrins: protoporphyrins IX, III, and XIII. *J. Org. Chem.* **51**, 666–671 (1986).
52. Engels, W. R. Contributing software to the internet: the amplify program. *Trends Biochem. Sci.* **18**, 448–450 (1993).
53. Stemmer, W. P. C., Cramer, A., Ha, K. D., Brennan, T. M. & Heyneker, H. L. Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene* **164**, 49–53 (1995).
54. Moore, J. T., Uppal, A., Maley, F. & Maley, G. F. Overcoming inclusion body formation in a high-level expression system. *Protein Expr. Purif.* **4**, 160–163 (1993).

## LETTERS

# Femtosecond characterization of vibrational optical activity of chiral molecules

Hanju Rhee<sup>1,2</sup>, Young-Gun June<sup>1</sup>, Jang-Soo Lee<sup>1,2</sup>, Kyung-Koo Lee<sup>1,2</sup>, Jeong-Hyon Ha<sup>3</sup>, Zee Hwan Kim<sup>1</sup>, Seung-Joon Jeon<sup>1,3</sup> & Minhaeng Cho<sup>1,2,3</sup>

Optical activity<sup>1–3</sup> is the result of chiral molecules interacting differently with left versus right circularly polarized light. Because of this intrinsic link to molecular structure, the determination of optical activity through circular dichroism (CD) spectroscopy has long served as a routine method for obtaining structural information about chemical and biological systems in condensed phases<sup>4–6</sup>. A recent development is time-resolved CD spectroscopy, which can in principle map the structural changes associated with biomolecular function<sup>7</sup> and thus lead to mechanistic insights into fundamental biological processes. But implementing time-resolved CD measurements is experimentally challenging because CD is a notoriously weak effect (a factor of  $10^{-4}$ – $10^{-6}$  smaller than absorption). In fact, this problem has so far prevented time-resolved vibrational CD experiments. Here we show that vibrational CD spectroscopy with femtosecond time resolution can be realized when using heterodyned spectral interferometry to detect<sup>8–10</sup> the phase and amplitude of the infrared optical activity free-induction-decay field in time (much like in a pulsed NMR experiment). We show that we can detect extremely weak signals in the presence of large achiral background contributions, by simultaneously measuring with a femtosecond laser pulse the vibrational CD and optical rotatory dispersion spectra of dissolved chiral limonene molecules. We have so far only targeted molecules in equilibrium, but it would be straightforward to extend the method for the observation of ultrafast structural changes such as those occurring during protein folding or asymmetric chemical reactions. That is, we should now be in a position to produce ‘molecular motion pictures’<sup>11</sup> of fundamental molecular processes from a chiral perspective.

CD spectroscopy is considered an incisive tool for determining the secondary structure of proteins in solution<sup>4–6,12,13</sup>. But biomolecules participating in chemical or biological reactions often undergo ultrafast structural changes that modulate their chiro-optical properties. The desire to map the chirality changes of such reactive systems over time, as a means to gain insight into the underlying structural changes, thus naturally led to the development of transient electronic CD measurement techniques<sup>7,14–16</sup>. One approach to improve the relatively poor time resolution of electronic CD spectrometry and overcome the weak-signal problems uses an ellipsometric detection scheme<sup>7,14</sup>, but its time resolution limit is set by the (limited) speed of the measurement electronics, such as amplifying gates or transient digitizers. Another approach combines an ultrashort laser with an electro-optic modulator or a Babinet-Soleil compensator to measure electronic CD or optical rotatory dispersion (ORD)<sup>15,16</sup> signals with subpicosecond time resolution. These approaches should in principle enable one to monitor time-resolved CD spectra by first optically triggering the chemical or biochemical reaction, and then measuring

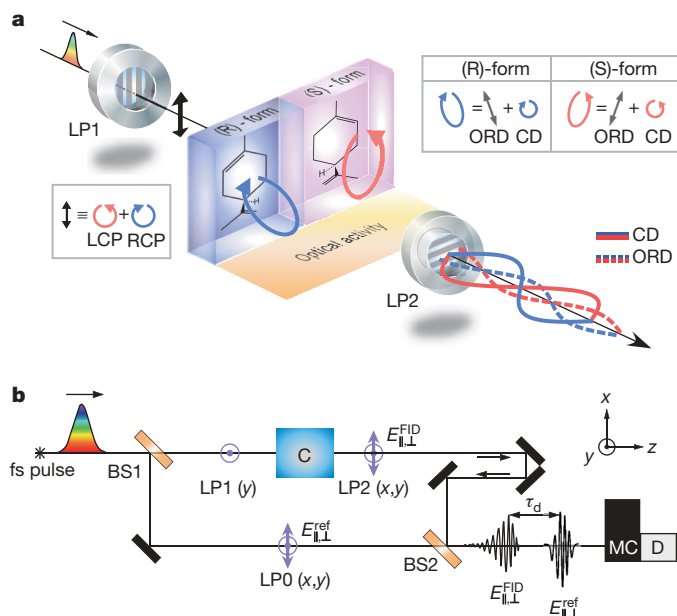
the differential absorption  $\Delta A$  of time-delayed light pulses that are left- and right-circularly polarized (LCP and RCP) or left- and right-elliptically polarized (LEP and REP) (that is,  $\Delta A = A_{\text{LCP}} - A_{\text{RCP}}$  or  $\Delta A = A_{\text{LEP}} - A_{\text{REP}}$ ). But the mode-locked ultrafast lasers used in such pump-probe measurements usually lack the intensity stabilities of approximately 0.001% that are needed to discriminate minuscule CD signals from the large achiral background signal.

Vibrational circular dichroism (VCD) is the vibrational analogue of the more popular and extensively used electronic CD. It is useful for differentiating between discrete structural conformations of biomolecules<sup>3,12,13,17</sup> and for determining absolute configurations of chiral molecules and drugs. Time-resolved VCD spectrometry could thus play a key role in unravelling the mechanisms of important biological or chemical processes, by monitoring the structural evolution of biomolecules or chiral molecules during reaction. But time-resolved VCD experiments are considered even more challenging than electronic CD experiments, mainly because the chiral susceptibilities for nuclear vibrations are much smaller than those for electronic degrees of freedom; consequently, neither an experimental set-up nor actual measurements involving ultrafast time-resolved VCD have been reported to date. Our approach to realizing femtosecond vibrational optical activity (CD and ORD) measurements involves detecting the phase and amplitude of the coherently emitted infrared optical activity free-induction-decay (FID) field (Fig. 1) using a novel cross-polarization detection scheme (Fig. 1a).

Figure 1a sketches the key principles of our experiment. Molecules with opposite chiral properties absorb LCP and RCP light differently. Therefore, an excitation pulse of linearly polarized light—an equal superposition of LCP and RCP—is transformed into LEP or REP light on interacting with chiral molecules. The interaction with the chiral molecules will at the same time give rise to circular birefringence, so that RCP light is transmitted with a different velocity from LCP light: the molecules rotate the major axis of the polarization ellipse arising from the CD phenomenon by an amount given by their circular birefringence  $\Delta n$ , defined as the differential index of refraction of LCP and RCP beams ( $\Delta n = n_{\text{LCP}} - n_{\text{RCP}}$ ). Particularly relevant here is that in the transmitted elliptically polarized light, information on the optical activity and hence chiral properties of the molecules is almost exclusively contained in the minor-axis field component (horizontal electric field component in Fig. 1a,  $\vec{E}_{\perp}$ ); in contrast, the major-axis field component (vertical electric field component,  $\vec{E}_{\parallel}$ ) is mostly influenced by all-electric-dipole-allowed achiral features of the molecule. As a result, a polarizer oriented perpendicular to the polarization direction of the excitation beam will select only the chiral (CD and ORD) components of the transmitted elliptically polarized light. It is the phase of the minor-axis field ( $\vec{E}_{\perp}$ ) that provides information on the handedness of the chiral molecule<sup>18</sup>. We found that the

<sup>1</sup>Department of Chemistry, <sup>2</sup>Center for Multidimensional Spectroscopy, Korea University, Seoul 136-701, Korea. <sup>3</sup>Multidimensional Spectroscopy Laboratory, Korea Basic Science Institute, Seoul 136-713, Korea.





**Figure 1 | Heterodyned detection of optical activity free induction decay.** **a**, Sketch of the basic principles of the cross-polarization detection in our optical activity FID measurements. We first linearly polarize a coherent femtosecond light pulse with linear polarizer LP1. The pulse is then incident on the chiral sample, and the left- and right-circularly polarized components (LCP, RCP) of the linearly polarized field are differentially absorbed. Depending on the chirality of the sample, this differential absorption (circular dichroism, CD) transforms the pulse into one with an eccentric left- or right-elliptical polarization (LEP, REP). Furthermore, the circular birefringence of the chiral molecules causes an optical rotation of the major (vertical) axis of the field polarization ellipse created through CD. After passage through the sample, the light pulse passes through linear polarizer LP2. LP2 is oriented perpendicular to LP1, so most of the transmitted field parallel to the incident beam is rejected while the perpendicular field components are retained. These perpendicular components are out of phase with each other after interaction of the light with two different enantiomers, the (R)-form (blue) and (S)-form (red); the individual CD (solid lines) and ORD (dashed lines) components are approximately in quadrature if the optical rotation angle is small. **b**, Sketch of the set-up for Fourier transform spectral interferometry experiments. The input pulse is split in two by a beam splitter (BS1). One beam is linearly y-polarized by LP1 and used to create electronic or vibrational coherences in the chiral sample (C), while the other is used to amplify and characterize the emitted FID field. For measurements of both parallel and perpendicular components of the FID fields ( $E_{\perp}^{FID}$  and  $E_{\parallel}^{FID}$ ) emitted from C, the transmission axes of the linear polarizers LP0 and LP2 are aligned along the y- and x-axis, respectively. The reference pulse ( $E_{\parallel}^{ref}$ ) and the signal field are combined by another beam splitter (BS2), with the former preceding the latter by a finite time,  $\tau_d$ . The heterodyned signal field is dispersed by a monochromator (MC) and detected by detector D.

relationship between  $\tilde{E}_{\perp}(\omega)$  and  $\tilde{E}_{\parallel}(\omega)$ , where  $\omega$  is frequency, is given by theory as (see Supplementary Information for details)

$$\tilde{E}_{\perp}(\omega) \propto \Delta\chi(\omega) \tilde{E}_{\parallel}(\omega) \quad (1)$$

where  $\tilde{E}_{\parallel}(\omega)$  is the complex transmitted electric field spectrum that results from the interference between the input field and the induced electric-dipole-allowed optical FID field. The complex function  $\Delta\chi(\omega)$  is the optical activity susceptibility; its imaginary part directly corresponds to the CD spectrum  $\Delta\alpha(\omega)$ , and its real part to the frequency-dependent ORD spectrum. Although the minor-axis component  $\tilde{E}_{\perp}(\omega)$  is extremely weak, the optical heterodyning allows us to amplify the weak signal field to give readily measurable levels of phase and amplitude information. Once the minor- and major-axis field components,  $\tilde{E}_{\perp}(\omega)$  and  $\tilde{E}_{\parallel}(\omega)$ , are measured, from equation (1) the complex susceptibility  $\Delta\chi(\omega)$  can be directly obtained.

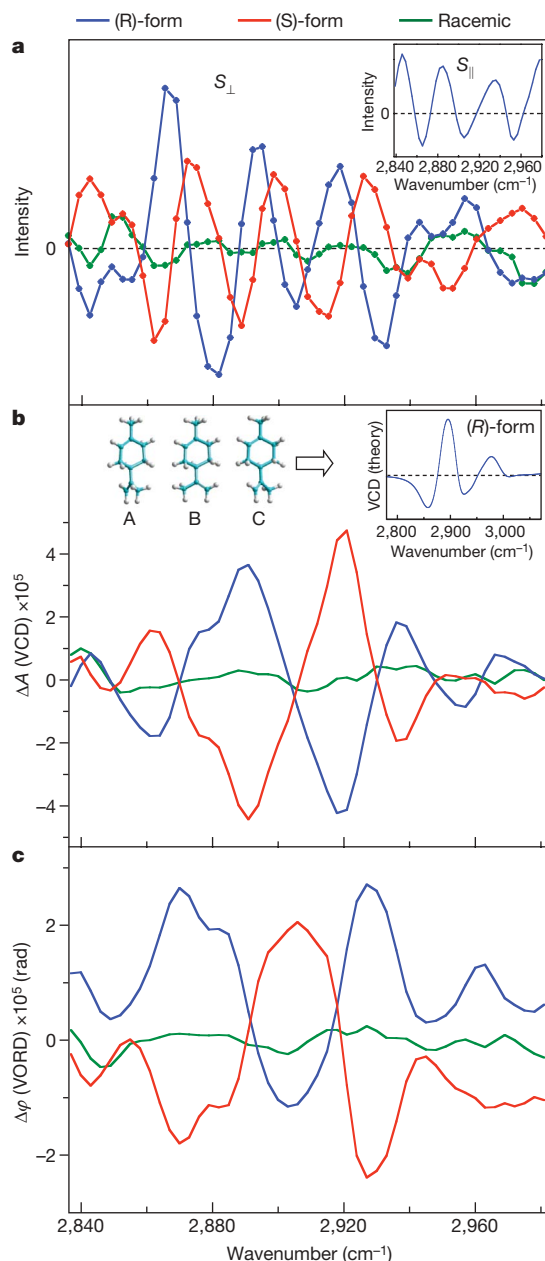
Our femtosecond measurement technique of vibrational optical activity utilizes heterodyned Fourier transform spectral

interferometry<sup>8–10,19–22</sup> as sketched in Fig. 1b (see Supplementary Information for more details regarding the experimental set-up). Briefly, we use an infrared pulse (1 kHz repetition rate, 2  $\mu$ J, centre wavelength of  $\sim 3.4 \mu$ m, and pulse width of  $\sim 60$  fs) generated by difference frequency mixing of a signal and an idler pulse from an optical parametric amplifier. To measure  $\tilde{E}_{\perp}(\omega)$ , the linear polarizers LP0 and LP2 are directed along the x-axis while the linear polarizer LP1 is oriented along the y-axis. In this cross-polarization configuration, only the  $\tilde{E}_{\perp}(\omega)$  component of the elliptically polarized light pulse that is transmitted through the sample is allowed to pass through LP2 for heterodyne detection. (A high-resolution motorized rotational stage is used to finely control the relative orientation of LP1 with incremental resolution of  $0.0005^\circ$ .) This FID signal and the reference pulse are then combined and spectrally dispersed with a monochromator, and the cross-polarization spectral interferogram  $S_{\perp}(\omega)$  is detected. To measure the parallel-polarization spectral interferogram  $S_{\parallel}(\omega)$  that originates from  $\tilde{E}_{\parallel}(\omega)$ , we rotate LP1 by a mere  $0.5^\circ$ ; this alters the beam pathway and experimental set-up only minimally, and the phase change induced by the rotated linear polarizer is thus negligibly small. We find that the performances of the polarizers (LP1 and LP2) are critical for the success of the experiment because  $\tilde{E}_{\perp}(\omega)$  is typically several orders of magnitude smaller than  $\tilde{E}_{\parallel}(\omega)$  when LP1  $\parallel$  LP2. We therefore chose calcite plate polarizers<sup>23</sup> for the main polarizer pair that removes the large achiral signal contribution  $\tilde{E}_{\parallel}(\omega)$ ; unlike usual prism-type polarizers (with extinction ratios of  $\sim 10^{-6}$ ), calcite plate polarizers utilize dichroic absorption that results in exceptionally small extinction ratios ( $10^{-8}$ – $10^{-9}$ ) within limited infrared spectral regions (3.35–3.5  $\mu$ m, 3.9–4.1  $\mu$ m, and so on).

To test both the cross-polarization detection scheme as a means for realizing femtosecond vibrational optical activity spectrometry (Fig. 1a) and our experimental set-up (Fig. 1b), we performed proof-of-principle measurements on the (R)- and (S)-enantiomers of limonene and their racemic mixture. Even though the working ranges of our calcite dichroic polarizers are limited, most of the C–H stretching mode frequencies of limonene fall within one of their spectral windows. Figure 2a shows the dispersed spectral interferograms (that is,  $S_{\perp}(\omega)$  and  $S_{\parallel}(\omega)$ ) measured in the 2,840 to 3,000  $\text{cm}^{-1}$  wavenumber range; all three spectral interferograms in Fig. 2a exhibit systematic differences in terms of oscillating amplitudes and phases. We note here that the heterodyne signal measured for the racemic mixture is finite (though relatively small) rather than zero; this signal is due to an optical imperfection of the linear polarizers, and originates from interference between the reference pulse and a weak horizontal light pulse component leaked from LP1. Although not ideal, this achiral measured signal component contributes only to the VCD spectrum as a background offset (Supplementary Fig. 3) and can thus be removed.

The CD and ORD spectra are obtained by transforming the measured spectral interferograms  $S_{\perp}(\omega)$  and  $S_{\parallel}(\omega)$  into the electric field spectra  $\tilde{E}_{\perp}(\omega)$  and  $\tilde{E}_{\parallel}(\omega)$  and then taking the ratio  $\tilde{E}_{\perp}(\omega)/\tilde{E}_{\parallel}(\omega)$ ; the imaginary and real parts of this ratio then yield the CD and ORD spectra, respectively. The spectral characteristics (amplitude and phase) of the reference pulse and an additional phase term produced by the time delay  $\tau_d$  between the signal field and the reference pulse (Fig. 1b) contribute equally to  $S_{\perp}(\omega)$  and  $S_{\parallel}(\omega)$ , so their contributions automatically cancel out when taking the ratio of  $\tilde{E}_{\perp}(\omega)$  to  $\tilde{E}_{\parallel}(\omega)$ . This means that as long as the two paths in the interferometer have a constant delay time  $\tau_d$  and are sufficiently stable (in terms of phase fluctuations) during measurement, CD and ORD spectra can be retrieved from the measured interferograms without requiring a detailed characterization of the reference pulse or precise determination of the value of the delay time  $\tau_d$ . This feature is crucial for significantly boosting the sensitivity of our method.

Figure 2b depicts VCD spectra  $\Delta\alpha(\omega)$  extracted from the spectral interferograms measured with samples of (R)- and (S)-limonene and their racemic mixture. The VCD peak positions of (R)- and (S)-limonene are identical and their signs opposite to each other, and



**Figure 2 | Vibrational optical activity signals of chiral limonenes.** **a**, Cross-polarization spectral interferograms  $S_{\perp}(\omega)$  of (*R*)-limonene (blue line) and (*S*)-limonene (red line) and their 1:1 racemic solution (green line) dissolved in  $\text{CCl}_4$ . (Analyte concentration, 1.2–1.5 M; path length,  $\sim 50 \mu\text{m}$ ; maximum absorbance, 1.1 for (*R*)-limonene and racemic solution, and 1.4 for (*S*)-limonene.) Parallel-polarization spectral interferograms  $S_{\parallel}(\omega)$  are required to retrieve VCD and VORD spectra, and the inset therefore gives that of (*R*)-limonene. Coloured dots give the measured data points for comparison, with the dashed line indicating zero. Note that the three limonene samples have distinctive spectral phases and amplitudes, which are different from one another. **b**, VCD spectra retrieved from the measured spectral interferograms in **a**, with linear baseline correction to enable a clear comparison. Absolute VCD values ( $\Delta A$ ) were directly calculated according to  $(4/2.303)\text{Im}[\tilde{E}_{\perp}(\omega)/\tilde{E}_{\parallel}(\omega)]$  (Supplementary Information). *Ab initio* calculations using MP2/6-311++G\*\* identified the three major (*R*)-limonene conformers A, B and C (top-left inset) and gave the simulated C–H stretch VCD spectrum shown in the top-right inset. **c**, VORD spectra extracted from the real part of  $\Delta\chi(\omega)$ . The absolute VORD values ( $\Delta\phi$ ) were also calculated by  $\text{Re}[\tilde{E}_{\perp}(\omega)/\tilde{E}_{\parallel}(\omega)]$  (Supplementary Information), and the maximum optical rotation angle (at  $2,927 \text{ cm}^{-1}$ ) found to be about  $1.6 \times 10^{-3}$  degrees.

the spectra are quantitatively consistent with spectra measured using a continuous wave Fourier transform infrared (FT-IR) VCD spectrometer<sup>24</sup> (Supplementary Fig. 5). This indicates that our technique is robust and reliable. The racemic spectrum does not show any notable structure, as expected. A series of concentration-dependent VCD measurements confirmed that the attenuation of the signal field by self-absorption does not affect the spectrum as long as the experimental optical density is close to or less than 1. We note that the data collection time needed to obtain the VCD spectra in Fig. 2b is only tens of minutes, whereas measurements with the commercial FT-IR VCD spectrometer required multiple hours of signal averaging. The enhanced sensitivity of our method is due to the fact that the measurements are non-differential, free from achiral background signal, and based on heterodyned detection.

Finally, Fig. 2c shows the vibrational ORD (VORD) spectra of the (*R*)- and (*S*)-limonene and their racemic mixture obtained directly from the real part of  $\Delta\chi(\omega)$ . The spectra illustrate successful location of a maximum in the angle of optical rotation of an infrared beam by (*R*)-limonene in solution that is as small as about  $10^{-5}$  rad. We note that resonant VORD spectra of a liquid crystal<sup>25</sup> and off-resonant VORD spectra of polypeptides<sup>26</sup> have been measured before; but the direct measurement of the resonant VORD spectrum of a small chiral molecule in solution presumably posed significant experimental challenges due to the minute extent of optical rotation and the large attenuation of the signal in the resonant frequency region. To enable a comparison between theory and experiment, we carried out high-level *ab initio* calculations of an isolated (*R*)-limonene molecule and quantum mechanical/molecular mechanical simulations of limonene/ $\text{CCl}_4$  solutions. The calculations identified three major conformers (Fig. 2b top left inset) and yielded a simulated VCD spectrum (Fig. 2b top right inset) in good agreement with experiment; this illustrates that the current quantum chemistry calculation method, when properly combined with molecular dynamics simulation techniques, is able to provide a quantitative description of the absolute configuration of a given chiral molecule in solution<sup>27–29</sup>.

Although we have not yet carried out time-resolved spectroscopy, the present work clearly establishes that femtosecond optical pulses can be used to simultaneously measure VCD and VORD spectra with a time resolution that is only limited by the optical activity FID time. Time-resolved CD (ORD) experiments using the current technique can therefore be implemented by adding a pump pulse to trigger chemical or physical changes (such as protein unfolding) and then using a femtosecond optical pulse to probe the subsequent relaxation process after varying waiting times,  $T_w$ . This would result in a series of VCD and VORD spectra as a function of  $T_w$ , and provide information on the conformational change of the target molecule. In most experiments so far, it was found that the pump-induced changes result in signals that are too weak to be discriminated from large static signals when trying to differentiate measured CD spectra at different times; thus, a pump modulation technique that allows one to measure only the pump-induced changes would be required to increase sensitivity and enable measurements. In contrast to previous time-resolved CD experiments relying on differential measurements, the present method does not require any polarization modulation of the probe pulse and should thus enable the development of a highly sensitive time-resolved CD version in a straightforward manner.

Our phase-and-amplitude measurement technique overcomes both the small-signal and non-zero-background problems of conventional electronic and vibrational CD techniques relying on differential absorption measurement, owing to its capability of coherent amplification of a weak signal field. Although the present work demonstrated infrared CD and ORD measurements, the key concepts and the technique are sufficiently general that they can be readily applied to the measurement of other related chiro-optical properties. We therefore anticipate that the present method, which is a CD analogue of pulsed NMR, will trigger novel developments to

further improve time-resolved optical activity spectroscopy for use in studies of biomolecular dynamics in aqueous solution at physiological conditions.

Received 7 April 2008; accepted 20 January 2009.

- Rosenfeld, L. Quantenmechanische theorie der natürlichen optischen aktivität von flüssigkeiten und gasen. *Z. Phys.* **52**, 161–174 (1928).
- Gillard, R. D. Hand in hand. *Nature* **347**, 594 (1990).
- Nakanishi, K., Berova, N. & Woody, R. W. *Circular Dichroism: Principles and Applications* (Wiley-VCH, 2000).
- Tinoco, I. Jr, Mickols, W., Maestre, M. F. & Bustamante, C. Absorption, scattering, and imaging of biomolecular structures with polarized light. *Annu. Rev. Biophys. Biophys. Chem.* **16**, 319–349 (1987).
- Chin, D.-H., Woody, R. W., Rohl, C. A. & Baldwin, R. L. Circular dichroism spectra of short, fixed-nucleus alanine helices. *Proc. Natl Acad. Sci. USA* **99**, 15416–15421 (2002).
- Clarke, D. T., Doig, A. J., Stapley, B. J. & Jones, G. R. The  $\alpha$ -helix folds on the millisecond time scale. *Proc. Natl Acad. Sci. USA* **96**, 7232–7237 (1999).
- Goldbeck, R. A., Kim-Shapiro, D. B. & Kliger, D. S. Fast natural and magnetic circular dichroism spectroscopy. *Annu. Rev. Phys. Chem.* **48**, 453–479 (1997).
- Lepetit, L., Chériaux, G. & Joffre, M. Linear techniques of phase measurement by femtosecond spectral interferometry for applications in spectroscopy. *J. Opt. Soc. Am. B* **12**, 2467–2474 (1995).
- Brixner, T. *et al.* Two-dimensional spectroscopy of electronic couplings in photosynthesis. *Nature* **434**, 625–628 (2005).
- Zanni, M. T., Ge, N.-H., Kim, Y. S. & Hochstrasser, R. M. Two-dimensional IR spectroscopy can be designed to eliminate the diagonal peaks and expose only the cross peaks needed for structure determination. *Proc. Natl Acad. Sci. USA* **98**, 11265–11270 (2001).
- Cho, M. Spectroscopy: Molecular motion pictures. *Nature* **444**, 431–432 (2006).
- Keiderling, T. A. Vibrational CD of biopolymers. *Nature* **322**, 851–852 (1986).
- Silva, R. A. G. D., Kubelka, J., Bour, P., Decatur, S. M. & Keiderling, T. A. Site-specific conformational determination in thermal unfolding studies of helical peptides using vibrational circular dichroism with isotopic substitution. *Proc. Natl Acad. Sci. USA* **97**, 8318–8323 (2000).
- Lewis, J. W. *et al.* New technique for measuring circular dichroism changes on a nanosecond time scale. Application to (carbonmonoxy)myoglobin and (carbonmonoxy)hemoglobin. *J. Phys. Chem.* **89**, 289–294 (1985).
- Xie, X. & Simon, J. D. Picosecond time-resolved circular dichroism study of protein relaxation in myoglobin following photodissociation of carbon monoxide. *J. Am. Chem. Soc.* **112**, 7802–7803 (1990).
- Nieborala, C. & Hache, F. Measuring the dynamics of circular dichroism in a pump-probe experiment with a Babinet-Soleil compensator. *J. Opt. Soc. Am. B* **23**, 2418–2424 (2006).
- Choi, J. H. & Cho, M. Amide I vibrational circular dichroism of dipeptide: Conformation dependence and fragment analysis. *J. Chem. Phys.* **120**, 4383–4392 (2004).
- Rhee, H., Ha, J.-H., Jeon, S.-J. & Cho, M. Femtosecond spectral interferometry of optical activity: Theory. *J. Chem. Phys.* **129**, 135102 (2008).
- Walecki, W. J., Fittinghoff, D. N., Smirl, A. L. & Trebino, R. Characterization of the polarization state of weak ultrashort coherent signals by dual-channel spectral interferometry. *Opt. Lett.* **22**, 81–83 (1997).
- Gallagher, S. M. *et al.* Heterodyne detection of the complete electric field of femtosecond four-wave mixing signals. *J. Opt. Soc. Am. B* **15**, 2338–2345 (1998).
- Brixner, T., Stiopkin, I. V. & Fleming, G. R. Tunable two-dimensional femtosecond spectroscopy. *Opt. Lett.* **29**, 884–886 (2004).
- Cowan, M. L. *et al.* Ultrafast memory loss and energy redistribution in the hydrogen bond network of liquid H<sub>2</sub>O. *Nature* **434**, 199–202 (2005).
- Bridges, T. J. & Klüber, J. W. Dichroic calcite polarizers for the infrared. *Appl. Opt.* **4**, 1121–1125 (1965).
- Guo, C. *et al.* Fourier transform vibrational circular dichroism from 800 to 10,000 cm<sup>-1</sup>: near-IR-VCD spectral standards for terpenes and related molecules. *Vib. Spectrosc.* **42**, 254–272 (2006).
- Schrader, V. B. & Korte, E. H. Infrarot-Rotationsdispersion (IRD). *Angew. Chem.* **84**, 218–219 (1972).
- Chirgadze, Yu. N., Venyaminov, S. Yu. & Lobachev, V. M. Optical rotatory dispersion of polypeptides in the near-infrared region. *Biopolymers* **10**, 809–820 (1971).
- Choi, J.-H., Hahn, S. & Cho, M. Vibrational spectroscopic characteristics of secondary structure polypeptides in liquid water: Constrained MD simulation studies. *Biopolymers* **83**, 519–536 (2006).
- Choi, J.-H., Cheon, S., Lee, H. & Cho, M. Two-dimensional nonlinear optical activity spectroscopy of coupled multi-chromophore system. *Phys. Chem. Chem. Phys.* **10**, 3839–3856 (2008).
- Cho, M. Coherent two-dimensional optical spectroscopy. *Chem. Rev.* **108**, 1331–1418 (2008).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This work was supported by Creative Research Initiatives (CMDS) of MEST/KOSEF (M.C.) and the Frontier Research Laboratory Program of KBSI (S.-J.J.).

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to M.C. (mcho@korea.ac.kr).



## LETTERS

# Hybrid organic–inorganic rotaxanes and molecular shuttles

Chin-Fa Lee<sup>1</sup>, David A. Leigh<sup>1</sup>, Robin G. Pritchard<sup>2</sup>, David Schultz<sup>1</sup>, Simon J. Teat<sup>3</sup>, Grigore A. Timco<sup>2</sup> & Richard E. P. Winpenny<sup>2</sup>

The tetravalency of carbon and its ability to form covalent bonds with itself and other elements enables large organic molecules with complex structures, functions and dynamics to be constructed. The varied electronic configurations and bonding patterns of inorganic elements, on the other hand, can impart diverse electronic, magnetic, catalytic and other useful properties to molecular-level structures. Some hybrid organic–inorganic materials that combine features of both chemistries have been developed, most notably metal–organic frameworks<sup>1</sup>, dense and extended organic–inorganic frameworks<sup>2</sup> and coordination polymers<sup>3</sup>. Metal ions have also been incorporated into molecules that contain interlocked subunits, such as rotaxanes<sup>4–7</sup> and catenanes<sup>6,8</sup>, and structures in which many inorganic clusters encircle polymer chains have been described<sup>9</sup>. Here we report the synthesis of a series of discrete rotaxane molecules in which inorganic and organic structural units are linked together mechanically at the molecular level. Structural units (dialkylammonium groups) in dumb-bell-shaped organic molecules template the assembly of essentially inorganic ‘rings’ about ‘axles’ to form rotaxanes consisting of various numbers of rings and axles. One of the rotaxanes behaves as a ‘molecular shuttle’<sup>10</sup>: the ring moves between two binding sites on the axle in a large-amplitude motion typical of some synthetic molecular machine systems<sup>11–15</sup>. The architecture of the rotaxanes ensures that the electronic, magnetic and paramagnetic characteristics of the inorganic rings—properties that could make them suitable as qubits for quantum computers<sup>16–18</sup>—can influence, and potentially be influenced by, the organic portion of the molecule.

The basis for the hybrid organic–inorganic rotaxane synthesis lies in the observation<sup>19</sup> that the formation of heterometallic rings of various shapes and sizes, containing seven or more trivalent Cr(III) ions and one or two divalent metal ions (typically Ni(II), Co(II), Fe(II) or Cu(II)) bridged by multiple fluoride and alkyl/aryl carboxylate anions, is templated by various organic cations<sup>20</sup>, including imidazolium<sup>21</sup>, *N*-alkylimidazolium<sup>21</sup> and primary<sup>22</sup> and secondary<sup>19</sup> ammonium groups. Dialkylammonium salts have previously been used to direct the assembly of rotaxanes based on crown ethers<sup>23,24</sup>, cucurbituril<sup>25,26</sup> and cyclic peptides<sup>27</sup>, so it seemed possible that they could also template the formation of rotaxanes using heterometallic rings. Secondary amine threads (1a–c) were constructed with bulky ‘stoppers’ at each end of the axle to prevent subsequent de-threading of the heterometallic ring assembled around the ammonium template. The threads were reacted<sup>19</sup> with a 7:1 molar ratio of Cr(III):Co(II) ions in the form of chromium(III) fluoride (CrF<sub>3</sub>·4H<sub>2</sub>O) and a cobalt(II) pivalate salt ([Co<sub>2</sub>(H<sub>2</sub>O)(*t*-BuCO<sub>2</sub>)<sub>4</sub>(*t*-BuCO<sub>2</sub>H)<sub>4</sub>]; *t*-Bu, tert-butyl), with pivalic acid as the solvent, at 140 °C for 12 h (Fig. 1a). Although no rotaxane was observed with the shorter threads (1a and 1b), probably owing to there being insufficient space between the stoppers

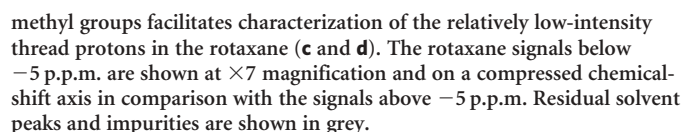
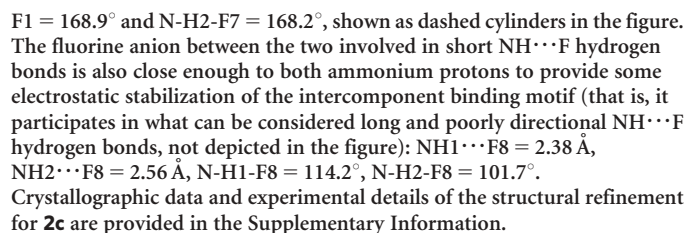
to accommodate the bulky heterometallic ring, we found that with a six-carbon-atom spacer between the ammonium centre and each stopper (thread 1c), [2]rotaxane 2c was produced in 23% yield, which is rather efficient for what is essentially a 33-component assembly process (each metal–ligand bond is dynamic under the reaction conditions).

Confirmation of the interlocked nature of 2c was provided by the solid-state structure of a single crystal of the [2]rotaxane grown from a saturated diethyl ether/acetonitrile solution (Fig. 1b). The [2]rotaxane is a neutral molecule, the positive charge on the ammonium thread being balanced by the charge remaining on a heterometallic macrocycle consisting of 24 monoanionic ligands (eight fluorides and sixteen pivalate groups) bound to seven trivalent cations (Cr(III)) and one divalent (Co(II)) cation. The crystal structure shows the ammonium cation in the centre of the cavity of the macrocycle, forming short (~2.0 Å) F<sup>−</sup>⋯HN<sup>+</sup> hydrogen bonds with two of the bridging fluoride ligands. Mass spectrometry confirms that each rotaxane molecule contains exactly one cobalt and seven chromium ions (Supplementary Information), the crystallographic data indicating that the cobalt ion is disordered over the eight different metal sites in each ring in the solid state (<sup>1</sup>H NMR suggests that it also has no particularly favoured position in solution (see below)).

The X-ray structure shows the eight metal centres in each heterometallic wheel to be almost perfectly coplanar, with the two faces of the ring almost identical in terms of physical shape and geometry. Nevertheless, the wheel is intrinsically chiral (that is, has a non-superimposable mirror image) because the one Co(II) ion must have either a (Δ) arrangement or a (Λ) arrangement of the ligands in its octahedral coordination sphere (see Supplementary Information for an explanation of this nomenclature). The chirality is particularly apparent in the unequal magnetic effect exerted on each geminal proton in the methylene groups of the thread as revealed by <sup>1</sup>H NMR spectroscopy (see below). The sixteen pivalate groups form a lipophilic outer coating for the heterometallic ring, conferring good solubility of the rotaxane in a range of organic solvents, including toluene, dichloromethane, chloroform and diethyl ether.

Comparison of the <sup>1</sup>H NMR spectrum of the parent thread, 1c (Fig. 2a), with that of the rotaxane, 2c (Fig. 2b), and a partially deuterated analogue, d<sub>144</sub>-2c (Fig. 2c, d), shows that the basic structure found in the solid state is retained in C<sub>2</sub>D<sub>2</sub>Cl<sub>4</sub> solution and also gives some information regarding the dynamics of the relative motion of the mechanically interlocked components. As one of the eight metal ions (Co(II)) is different from the other seven (Cr(III)) in each ring, if the heterometallic ring spun about the axis of the thread rapidly on the NMR timescale then the sixteen pivalate groups should experience eight non-equivalent sets of chemical environments—four for axial and four for equatorial pivalate groups—as observed

<sup>1</sup>School of Chemistry, University of Edinburgh, The King's Buildings, West Mains Road, Edinburgh EH9 3JJ, UK. <sup>2</sup>School of Chemistry, University of Manchester, Oxford Road, Manchester M13 9PL, UK. <sup>3</sup>Advanced Light Source, Lawrence Berkeley Laboratory, 1 Cyclotron Road, MS2-400, Berkeley, California 94720, USA.

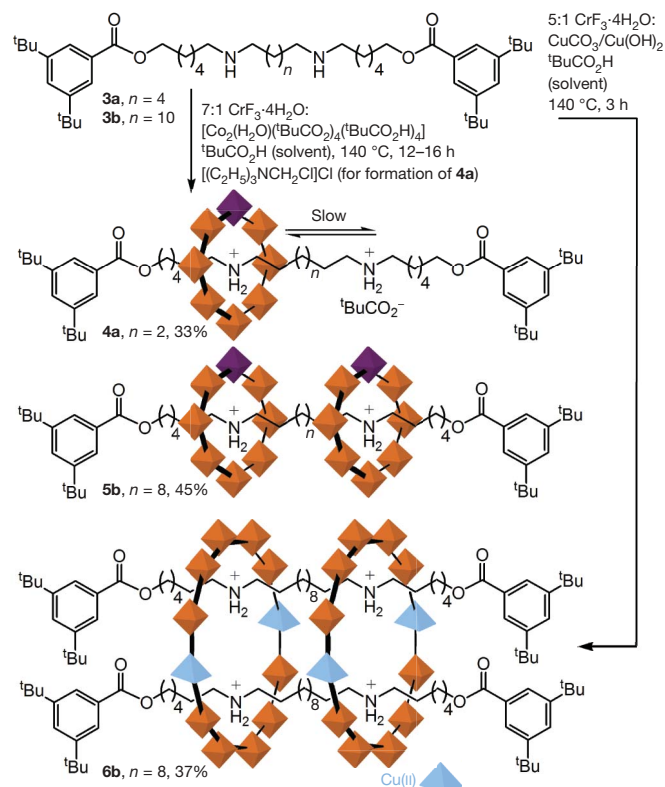


for the parent macrocycle  $[\text{Cr}_7\text{CoF}_8(t\text{-BuCO}_2)_{16}\cdot\text{H}_3\text{O}]$  (Supplementary Information). If, on the other hand, the ring were rotating slowly about the thread then it would become desymmetrized on the NMR timescale and many more non-equivalent sets of signals for the pivalate groups would be expected. Eight resonances for pivalate methyl groups (blue signals, Fig. 2b) are present for the [2]rotaxane at 330 K in  $\text{C}_2\text{D}_2\text{Cl}_4$  (500 MHz), and as cooling to 240 K does not change this number, nor appreciably broaden the signals, it appears that the heterometallic ring rotates about the thread (randomly under thermal motion) very rapidly under these conditions.

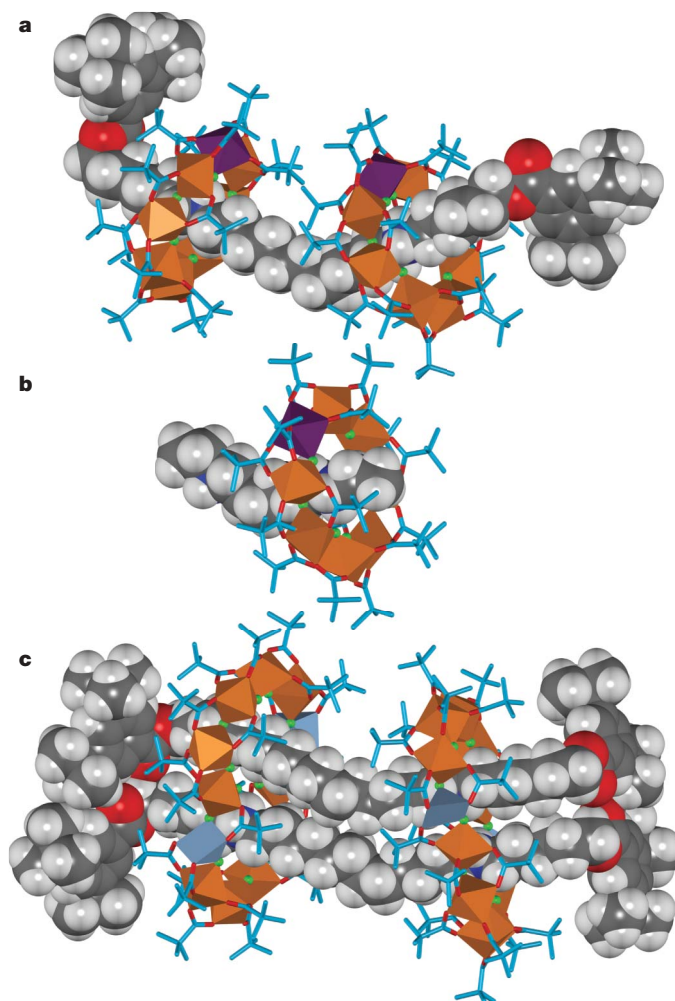
The most dramatic effects of interlocking the inorganic and organic components, however, are the very large shifts (up to 45 p.p.m.; Fig. 2d) in the thread protons caused by the paramagnetic  $\text{Co}(\text{II})$  ion, the greatest shifts generally occurring for the protons closest to the heterometallic wheel. The presence of two signals for each pair of methylene protons is due to the two protons of each methylene group (labelled primed and unprimed in Fig. 2) being diastereotopic (that is, magnetically distinct) given the chirality of the heterometallic wheel. The magnitude of the shifts is highly sensitive to changes in temperature (for example, there is a 6–8 p.p.m. change in the chemical shift of  $\text{H}_b$  and  $\text{H}_b'$  over a range of just 30 K; compare Fig. 2c with Fig. 2d).

The hybrid organic–inorganic rotaxane synthesis was successfully extended to threads containing more than one ammonium binding site (Fig. 3). This enabled the preparation of significantly more complex architectures (Fig. 4), including multi-ring [3]- and [4]rotaxanes and dynamic molecular shuttles in which the ring moves randomly under thermal motion between different binding sites on the thread.

With a long (12-carbon-atom) spacer between two ammonium groups on the thread (3b), reaction under the previously established conditions (7:1  $\text{CrF}_3\cdot 4\text{H}_2\text{O}:[\text{Co}_2(\text{H}_2\text{O})(t\text{-BuCO}_2)_4(t\text{-BuCO}_2\text{H})_4]$ ,  $t\text{-BuCO}_2\text{H}$ , 140 °C, 12 h) afforded [3]rotaxane 5b in 45% yield (from



**Figure 3 | Synthesis of hybrid organic–inorganic [3]rotaxane 5b, [4]rotaxane 6b and molecular shuttle 4a.** Compound 5b has two rings on one thread, 6b has two rings each circumscribing two threads and 4a is a [2]rotaxane in which the ring moves randomly between the two ammonium sites on the thread, restricted by the tight ion-pairing and intercomponent  $\text{NH}\cdots\text{F}$  hydrogen-bonding.



**Figure 4 | X-ray crystal structures.** a, [3]rotaxane 5b. b, A pseudorotaxane featuring the bisammonium binding-site arrangement present in molecular shuttle 4a. c, [4]rotaxane 6b. The atom representations are as indicated in the legend of Fig. 1b, except that here the organic threads are shown at van der Waals radii and the metal ions of the macrocycles are depicted by solid polyhedra reflecting their coordination geometries:  $\text{Cr}(\text{III})$ , orange octahedron;  $\text{Co}(\text{II})$ , purple octahedron;  $\text{Cu}(\text{II})$ , blue square pyramid. Details of each set of intercomponent binding motifs are provided in the Supplementary Information. In a and b the divalent metal ion,  $\text{Co}(\text{II})$ , is disordered over the eight metal sites on each heterometallic wheel (as for [2]rotaxane 2c; Fig. 1b); in c the divalent metal ion,  $\text{Cu}(\text{II})$ , is localized at the positions indicated. The crystallographic data and experimental details of the structural refinement for 4a, 5b and 6b are provided in the Supplementary Information.

65 components), without any significant quantity (<1%) of the corresponding [2]rotaxane. The X-ray structure (Fig. 4a) of a single crystal of 5b grown from a saturated solution in tetrahydrofuran/acetonitrile shows types of intercomponent interactions similar to those observed with [2]rotaxane 2c, although at one of the macrocycle–thread binding sites the four  $\text{NH}\cdots\text{F}$  contacts are of such similar lengths that they can be considered to be pairs of bifurcated  $\text{F}^-\cdots\text{H}(\text{N}^+\text{H})\cdots\text{F}^-$  hydrogen bonds (Supplementary Information).

Notably, changing the divalent metal from  $\text{Co}(\text{II})$  to  $\text{Cu}(\text{II})$  led to a [4]rotaxane, 6b, with two thread molecules circumscribed by two heterometallic macrocycles<sup>28</sup>, each ring containing ten octahedral  $\text{Cr}(\text{III})$  ions and two square-pyramidal  $\text{Cu}(\text{II})$  ions. By modifying the ratio of the trivalent:divalent metal-salt reagents from 7:1 to the 5:1 required for the stoichiometry of this product, we isolated [4]rotaxane 6b in 37% yield (from 98 components: 24 metal ions, 44 pivalate groups, 28 fluoride ions and two organic threads). The X-ray structure (Fig. 4c) of a single crystal of 6b grown from a saturated



solution of the [4]rotaxane in diethyl ether and acetonitrile was solved with data collected using the Advanced Light Source synchrotron, Lawrence Berkeley National Laboratory. The different coordination requirements of Cu(II) (five coordinate) compared with Co(II) (six coordinate) leads to some enforced changes in the ligand arrangement for the larger heterometallic wheel, and in the [4]rotaxane two of the 14 fluoride ligands are bound to only one metal ion rather than bridging two. These two singly coordinated fluoride anions have a central role as hydrogen-bond acceptors in the intercomponent binding motifs of the [4]rotaxane in the solid state, each being involved in bifurcated hydrogen-bonding to two ammonium groups on different threads (Supplementary Information).

By shortening the spacer between the ammonium units on the thread from twelve (**3b**) to six methylene groups (**3a**)—which is too short to allow heterometallic rings to reside simultaneously on both ammonium sites—and adding (chloromethyl)triethylammonium chloride to the reaction, we were able to form a [2]rotaxane with one heterometallic ring on a thread with two ammonium groups, **4a**, in 33% yield (Fig. 3). The (chloromethyl)triethylammonium chloride was originally introduced accidentally to the reaction as an impurity, produced by the alkylation of triethylamine by dichloromethane during isolation of the thread by column chromatography. However, when subsequent reactions using pure samples of the thread failed to produce any of rotaxane **4a**, analysis of the original procedure established the presence of (chloromethyl)triethylammonium chloride in the successful rotaxane-forming experiment. When (chloromethyl)triethylammonium chloride was deliberately added to the pure thread, the reaction once again produced rotaxane **4a**. Other tetraalkylammonium salts, such as tetraethylammonium chloride, did not promote rotaxane formation when added to the reaction mixture. This suggests that the chloromethyl group has a crucial role, possibly as an electrophilic site of temporary attachment for the tetraalkylammonium cation, which might then act as a macrocyclization template for a growing oligomeric metal-pivalate-fluoride chain.

Unusually for a hydrogen-bonded molecular shuttle, the random back-and-forth shuttling of the macrocycle between the two binding sites on the thread in **4a** occurs on a very different timescale from the spinning of the ring about the axis of the thread. At 300 K, the 500-MHz  $^1\text{H}$  NMR spectrum of **4a** in a range of solvents ( $\text{CDCl}_3$ ,  $\text{C}_2\text{D}_2\text{Cl}_4$ ,  $\text{d}_8$ -tetrahydrofuran) shows a well-resolved pair of signals for each chemically distinct 'stopper' proton (Supplementary Information), the difference within each pair being caused by whether the adjacent ammonium centre is vacant or occupied by the heterocyclic ring, indicating that the shuttling between them is slow on the NMR timescale. Heating to 360 K ( $\text{C}_2\text{D}_2\text{Cl}_4$  as solvent) did not significantly broaden, let alone coalesce, the signal pairs. However, the dynamic exchange of the ring between the two thread ammonium sites could be measured by polarization transfer between signal pairs during the mixing time in two-dimensional exchange spectroscopy experiments<sup>29</sup> (Supplementary Information). These measurements gave a rate of exchange of  $1.2 \pm 0.5 \text{ s}^{-1}$  at 330 K in  $\text{C}_2\text{D}_2\text{Cl}_4$ , corresponding to a free energy of activation ( $\Delta G^\ddagger$ ) for shuttling of  $19.3 \pm 0.2 \text{ kcal mol}^{-1}$ . This value is at least (and possibly much more than)  $10 \text{ kcal mol}^{-1}$  higher than the activation barrier to spinning of the ring.

This significant difference can be rationalized from the intercomponent  $\text{NH}\cdots\text{F}$  hydrogen-bonding seen in the various crystal structures (Figs 1b and 4). During spinning, the pseudo- $\text{C}_8$ -symmetry of the bridging fluoride ions in the wheel enables the  $\text{NH}\cdots\text{F}$  hydrogen bonds to be transferred from one fluoride ion to the next by a low-energy route that does not require significant breaking of an existing hydrogen bond before a new one starts to form, that is, conversion from a 'short'  $\text{NH}\cdots\text{F}$  hydrogen bond to a 'long' one through a bifurcated system. These motifs are all seen in the X-ray structures of the various rotaxanes. By contrast, any shuttling mechanism for **4a** must require complete breaking of all the intercomponent  $\text{NH}\cdots\text{F}$

hydrogen bonds before translocation of the wheel to the other ammonium binding site can occur.

The linking of organic and inorganic components at the molecular level by mechanical bonding imparts features of both chemistries (for example dynamics and magnetism) to the overall structure, and introduces new and intriguing behaviour through the combination of the two (for example the differences in the mechanisms of rotational and translational intercomponent motion, new non-covalent binding motifs, extreme temperature sensitivity of chemical shifts, novel types of template mechanism and so on). The provision of a simple and versatile synthetic route for linking inorganic and organic components by mechanical bonds at the molecular level should make it possible to create molecules with composite properties, including new behaviours and characteristics that may arise by combining features that have previously been solely the preserve of one type of chemistry or the other.

Received 16 November 2008; accepted 29 January 2009.

- Yaghi, O. M. *et al.* Reticular synthesis and the design of new materials. *Nature* **423**, 705–714 (2003).
- Cheetham, A. K. & Rao, C. N. R. There's room in the middle. *Science* **318**, 58–59 (2007).
- Kitagawa, S., Kitaura, R. & Noro, S.-i. Functional porous coordination polymers. *Angew. Chem. Int. Edn* **43**, 2334–2375 (2004).
- Ogino, H. Relatively high-yield syntheses of rotaxanes. Syntheses and properties of compounds consisting of cyclodextrins threaded by  $\alpha$ ,  $\omega$ -diaminoalkanes coordinated to cobalt(III) complexes. *J. Am. Chem. Soc.* **103**, 1303–1304 (1981).
- Batten, S. R. & Robson, R. Interpenetrating nets: ordered, periodic entanglement. *Angew. Chem. Int. Edn* **37**, 1460–1494 (1998).
- Sauvage, J.-P. & Dietrich-Buchecker, C. (eds). *Molecular Catenanes, Rotaxanes and Knots: A Journey through the World of Molecular Topology* (Wiley-VCH, 1999).
- Loeb, S. J. Metal organic rotaxane frameworks. *Chem. Commun.* 1511–1518 (2005).
- Fujita, M., Ibukuro, F., Hagihara, H. & Ogura, K. Quantitative self-assembly of a [2]catenane from two preformed molecular rings. *Nature* **367**, 720–723 (1994).
- Alam, M. A. *et al.* Directed 1D assembly of a ring-shaped inorganic nanocluster templated by an organic rigid-rod molecule. *Angew. Chem. Int. Edn* **47**, 2070–2073 (2008).
- Anelli, P. L., Spencer, N. & Stoddart, J. F. A molecular shuttle. *J. Am. Chem. Soc.* **113**, 5131–5133 (1991).
- Berná, J. *et al.* Macroscopic transport by synthetic molecular machines. *Nature Mater.* **4**, 704–710 (2005).
- Nguyen, T. *et al.* A reversible molecular valve. *Proc. Natl Acad. Sci. USA* **102**, 10029–10034 (2005).
- Liu, Y. *et al.* Linear artificial molecular muscles. *J. Am. Chem. Soc.* **127**, 9745–9759 (2005).
- Green, J. E. *et al.* A 160-kilobit molecular electronic memory patterned at  $10^{11}$  bits per square centimetre. *Nature* **445**, 414–417 (2007).
- Kay, E. R., Leigh, D. A. & Zerbetto, F. Synthetic molecular motors and mechanical machines. *Angew. Chem. Int. Edn* **46**, 72–191 (2007).
- Leuenberger, M. N. & Loss, D. Quantum computing in molecular magnets. *Nature* **410**, 789–793 (2001).
- Bertina, S. *et al.* Quantum oscillations in a molecular magnet. *Nature* **453**, 203–206 (2008).
- Winpenny, R. E. P. Quantum information processing using molecular nanomagnets as qubits. *Angew. Chem. Int. Edn* **47**, 7992–7994 (2008).
- Larsen, F. K. *et al.* Synthesis and characterization of heterometallic  $\{\text{Cr}_7\text{M}\}$  wheels. *Angew. Chem. Int. Edn* **42**, 101–105 (2003).
- Affronte, M., Carretta, S., Timco, G. A. & Winpenny, R. E. P. A ring cycle: studies of heterometallic wheels. *Chem. Commun.* 1789–1797 (2007).
- Timco, G. A. *et al.* Influencing the nuclearity and constitution of heterometallic rings via templates. *Chem. Commun.* 3649–3651 (2005).
- Cador, O. *et al.* The magnetic Möbius strip: synthesis, structure and magnetic studies of odd-numbered antiferromagnetically coupled wheels. *Angew. Chem. Int. Edn* **43**, 5196–5200 (2004).
- Kolchinski, A. G., Busch, D. H. & Alcock, N. W. Gaining control over molecular threading: benefits of second coordination sites and aqueous–organic interfaces in rotaxane synthesis. *J. Chem. Soc. Chem. Commun.* 1289–1291 (1995).
- Ashton, P. R. *et al.* Self-assembling [2]- and [3]rotaxanes from secondary dialkylammonium salts and crown ethers. *Chem. Eur. J.* **2**, 729–736 (1996).
- Mock, W. L., Irra, T. A., Wepsiec, J. P. & Adhia, M. Catalysis by cucurbituril. The significance of bound-substrate destabilization for induced triazole formation. *J. Org. Chem.* **54**, 5302–5308 (1989).
- Kim, K. Mechanically interlocked molecules incorporating cucurbituril and their supramolecular assemblies. *Chem. Soc. Rev.* **31**, 96–107 (2002).
- Aucagne, V., Leigh, D. A., Lock, J. S. & Thomson, A. R. Rotaxanes of cyclic peptides. *J. Am. Chem. Soc.* **128**, 1784–1785 (2006).
- Ashton, P. R. *et al.* Doubly encircled and double-stranded pseudorotaxanes. *Angew. Chem. Int. Edn Engl.* **34**, 1869–1871 (1995).

29. Perrin, C. L. & Dwyer, T. J. Application of two-dimensional NMR to kinetics of chemical exchange. *Chem. Rev.* **90**, 935–967 (1990).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank J. Bella for the exchange spectroscopy NMR experiments, W. Sun for assistance with the preparation of thread **1c** and the Engineering and Physical Sciences Research Council (EPSRC) National Mass Spectrometry Service Centre (Swansea, UK) for high-resolution mass spectrometry. This research was funded by the European Commission (through the NoE 'MAGMANet') and EPSRC. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the US Department of Energy under contract no. DE-AC02-05CH11231. D.S. is a Swiss National Science Foundation postdoctoral fellow. D.A.L. is an EPSRC Senior Research Fellow and holds a Royal Society Wolfson Research Merit Award.

**Author Contributions** C.-F.L., D.S. and G.A.T. carried out the synthesis and characterization studies, helped plan the experiments and participated in the preparation of the manuscript. R.G.P. and S.J.T. collected the X-ray data and solved the crystal structures. D.A.L. and R.E.P.W. helped plan the experiments and prepare the manuscript.

**Author Information** The crystallographic data and experimental details of the structural refinement for the X-ray crystal structures reported in this paper have been deposited at the Cambridge Crystallographic Data Centre, under deposition numbers CCDC 705132–CCDC 705135. These data can be obtained free of charge from the Cambridge Crystallographic Data Centre ([http://www.ccdc.cam.ac.uk/data\\_request/cif](http://www.ccdc.cam.ac.uk/data_request/cif)). Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to D.A.L. ([david.leigh@ed.ac.uk](mailto:david.leigh@ed.ac.uk)) or R.E.P.W. ([richard.winpenny@manchester.ac.uk](mailto:richard.winpenny@manchester.ac.uk)).

# Temperature-dependent thermal diffusivity of the Earth's crust and implications for magmatism

Alan G. Whittington<sup>1</sup>, Anne M. Hofmeister<sup>2</sup> & Peter I. Nabelek<sup>1</sup>

The thermal evolution of planetary crust and lithosphere is largely governed by the rate of heat transfer by conduction<sup>1–3</sup>. The governing physical properties are thermal diffusivity ( $\kappa$ ) and conductivity ( $k = \kappa\rho C_p$ ), where  $\rho$  denotes density and  $C_p$  denotes specific heat capacity at constant pressure. Although for crustal rocks both  $\kappa$  and  $k$  decrease above ambient temperature<sup>4,5</sup>, most thermal models of the Earth's lithosphere assume constant values for  $\kappa$  ( $\sim 1 \text{ mm}^2 \text{ s}^{-1}$ ) and/or  $k$  ( $\sim 3$  to  $5 \text{ W m}^{-1} \text{ K}^{-1}$ )<sup>6,7</sup> owing to the large experimental uncertainties associated with conventional contact methods at high temperatures. Recent advances in laser-flash analysis<sup>8,9</sup> permit accurate ( $\pm 2$  per cent) measurements on minerals and rocks to geologically relevant temperatures<sup>10</sup>. Here we provide data from laser-flash analysis for three different crustal rock types, showing that  $\kappa$  strongly decreases from  $1.5$ – $2.5 \text{ mm}^2 \text{ s}^{-1}$  at ambient conditions, approaching  $0.5 \text{ mm}^2 \text{ s}^{-1}$  at mid-crustal temperatures. The latter value is approximately half that commonly assumed, and hot middle to lower crust is therefore a much more effective thermal insulator than previously thought. Above the quartz  $\alpha$ – $\beta$  phase transition, crustal  $\kappa$  is nearly independent of temperature, and similar to that of mantle materials<sup>11</sup>. Calculated values of  $k$  indicate that its negative dependence on temperature is smaller than that of  $\kappa$ , owing to the increase of  $C_p$  with increasing temperature, but  $k$  also diminishes by 50 per cent from the surface to the quartz  $\alpha$ – $\beta$  transition. We present models of lithospheric thermal evolution during continental collision and demonstrate that the temperature dependence of  $\kappa$  and  $C_p$  leads to positive feedback between strain heating in shear zones and more efficient thermal insulation, removing the requirement for unusually high radiogenic heat production to achieve crustal melting temperatures. Positive feedback between heating, increased thermal insulation and partial melting is predicted to occur in many tectonic settings, and in both the crust and the mantle, facilitating crustal reworking and planetary differentiation<sup>12</sup>.

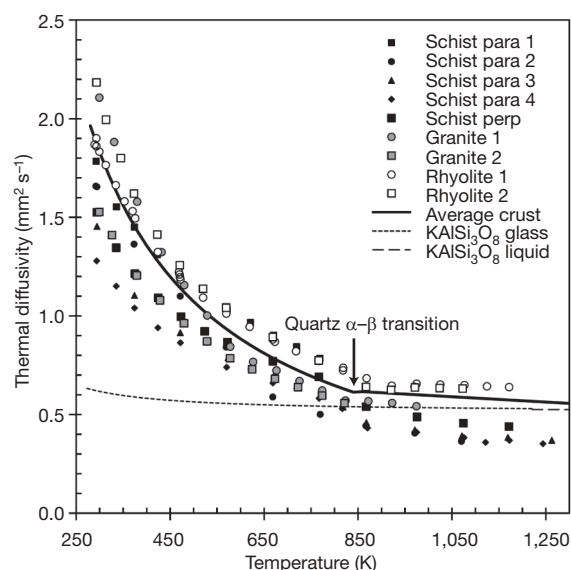
Thermal diffusivity data for garnet schist, leucogranite and welded rhyolitic ash-flow tuff were acquired at temperatures of up to  $1,260 \text{ K}$  using laser flash analysis, which isolates the phonon component of heat transfer from radiative transfer and avoids thermal contact losses<sup>13</sup>. Graphite-coated wafers approximately  $1\text{-mm}$  thick and  $10 \text{ mm}$  in diameter are held at some temperature in a furnace. Thermal diffusivity is determined from the time-dependent response of the sample to an additional pulse of heat, supplied by a gadolinium–gallium–garnet laser to the bottom of the sample, which is monitored using an indium antimonide detector above the sample. The leucogranite and rhyolite are homogeneous and isotropic. The schist is anisotropic owing to alternating mica- and quartz-rich layers, so samples were cut parallel and perpendicular to foliation.

Room-temperature ( $\sim 298 \text{ K}$ ) thermal diffusivities ranged from  $1.3$  to  $1.8 \text{ mm}^2 \text{ s}^{-1}$  for the schist, from  $1.5$  to  $2.1 \text{ mm}^2 \text{ s}^{-1}$  for the

granite and from  $1.8$  to  $2.2 \text{ mm}^2 \text{ s}^{-1}$  for the rhyolite (Fig. 1). For schist slices cut perpendicular to foliation, heat flow travels parallel to foliation and averages alternating compositional layers; these perpendicular slices yielded values that are intermediate between values for quartz-rich (high- $\kappa$ ) and mica-rich (low- $\kappa$ ) layers sampled by cuts parallel to foliation (when heat flow is across each layer). Variations in  $\kappa$  among different granite and rhyolite samples were also ascribed to differences in the amount of quartz sampled by the laser spot, which is  $6 \text{ mm}$  in diameter.

For all samples,  $\kappa$  decreased rapidly with increasing temperature, and asymptotically approached a high-temperature limit. Variations among rocks, and among different samples of the same rock, decreased markedly between room temperature and  $600 \text{ K}$ . Above  $\sim 750 \text{ K}$ , the  $\kappa$  of the schist samples became very low, probably as a result of differential thermal expansion and cracking along cleavage planes, perhaps accompanied by mica dehydration. The data suggest a break in slope at the quartz  $\alpha$ – $\beta$  transition temperature, followed by a slight, quasi-linear decrease in  $\kappa$  above  $846 \text{ K}$ . This trend is also observed in quartzites<sup>14</sup>.

Data for the two rhyolite samples, two granite samples and the perpendicular schist were fitted by the following equations, which



**Figure 1 | Thermal diffusivity of crustal rocks as a function of temperature.** Data are given in Supplementary Information. Uncertainties are smaller than symbol size in most cases. Curve for average crust calculated using equations (1) and (2) below and above the quartz  $\alpha$ – $\beta$  transition, respectively. Curves for  $\text{KAlSi}_3\text{O}_8$  glass and liquid are from ref. 28. para, parallel; perp, perpendicular.

<sup>1</sup>Department of Geological Sciences, University of Missouri, Columbia, Missouri 65211, USA. <sup>2</sup>Department of Earth and Planetary Sciences, Washington University, St Louis, Missouri 63130, USA.



should be a good estimate of the temperature dependence of thermal diffusivity in the continental crust:

$$\kappa_{\text{crust}}(T < 846\text{K}) = 567.3/T - 0.062 \quad (1)$$

$$\kappa_{\text{crust}}(T > 846\text{K}) = 0.732 - 0.000135T \quad (2)$$

Here  $\kappa$  is measured in square millimetres per second and  $T$  (temperature) is measured in kelvin. Crystalline garnet, pyroxene and olivine all have low average  $\kappa$  values in the range  $0.6\text{--}0.7\text{ mm}^2\text{ s}^{-1}$  at upper-mantle temperatures<sup>10,11,15</sup>, so any change in  $\kappa$  across the crust–mantle boundary is likely to be minor. Rather, it is within the middle and upper crust that  $\kappa$  changes significantly, by a factor of 3–5.

To calculate thermal conductivity, we derived the following ‘bulk crustal’ specific heat capacity equations from end-member mineral data<sup>16</sup>, on the basis of a mineralogy of 30% quartz, 60% albite, 5% phlogopite and 5% annite:

$$C_{p,\text{crust}}(T < 846\text{K}) = 199.50 + 0.0857T - 5.0 \times 10^{-6}T^{-2} \quad (3)$$

$$C_{p,\text{crust}}(T > 846\text{K}) = 229.32 + 0.0323T - 47.9 \times 10^{-6}T^{-2} \quad (4)$$

Here  $C_p$  is measured in joules per mole per kelvin, the average molar mass is  $221.78\text{ g mol}^{-1}$  and  $T$  is measured in kelvin. Although the precise values of  $C_p$  depend on modal mineralogy, calculated differences in  $C_p$  among several possible assemblages such as tonalite, granite and schist are  $<3\%$  at  $1,000\text{ K}$ . We assumed a constant density of  $2,700\text{ kg m}^{-3}$ , because the increase in density due to compression is partly off-set by the decrease due to higher temperatures, and these changes are small in comparison with changes in  $C_p$  and  $k$ .

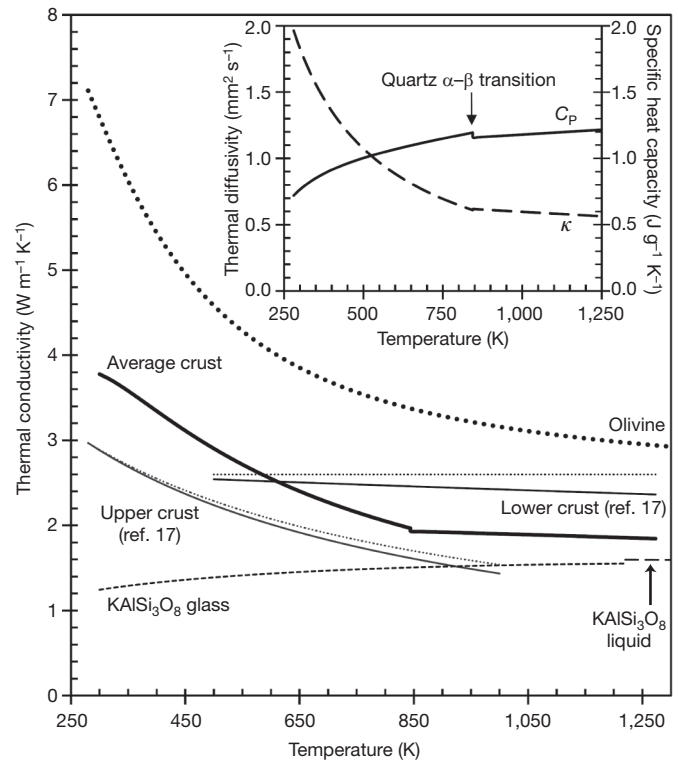
Calculated values for crustal  $k$  decrease with increasing temperature, from  $\sim 3.8\text{ W m}^{-1}\text{ K}^{-1}$  at the surface to  $\sim 1.9\text{ W m}^{-1}\text{ K}^{-1}$  at the quartz  $\alpha$ – $\beta$  transition, and then decrease only very slightly at higher temperatures (Fig. 2). Our calculations suggest that the  $k$  of the lower crust is 25% lower than previous estimates<sup>17</sup>, although this depends in part on assumed mineralogy. At  $1,000\text{ K}$ , the  $k$  of olivine<sup>15</sup> is  $3.0\text{ W m}^{-1}\text{ K}^{-1}$ , which is 60% higher than that of the crust. Therefore, the middle to lower crust is the most thermally insulating portion of continental lithosphere.

Negative temperature derivatives of  $\kappa$  and  $k$  may lead to a positive feedback between temperature increase and heat retention due to decreased thermal diffusivity, in the presence of a heat source. Potential heat sources for continental crust include heat flux from the mantle, radioactive decay, strain heating in deforming rocks and intrusion of mantle-derived basaltic magma. Collisional orogenic belts such as the Himalayas do not derive heat from basalts, yet crustal melting and the generation of leucogranites is a universal feature of such orogens<sup>18</sup>. The mechanism for achieving this has long been debated<sup>19</sup>, with most models favouring either unusually high radiogenic heat production in the middle crust<sup>20,21</sup> or strain heating in major shear zones<sup>22–24</sup>.

To test the effect of a temperature-dependent  $\kappa$  on the thermal structure of orogenic belts, we used the finite-difference code OROGEN<sup>25</sup> to model the one-dimensional thermal evolution of a doubly thickened crust. The code was modified to incorporate the temperature-dependent  $\kappa$  and  $C_p$  from equations (1), (2), (3) and (4), and solves the heat flow equation:

$$\frac{\partial T}{\partial t} = \frac{\partial}{\partial z} \left( \kappa \frac{\partial T}{\partial z} \right) + \frac{A_{\text{rad}} + A_s}{\rho C_p} \quad (5)$$

The initial conditions assume stacking of 35-km-thick crust giving the initial sawtooth geotherm. A vertical grid spacing of 250 m was used. The models included radiogenic heat production ( $A_{\text{rad}}$ ) as follows:  $A_{\text{rad}}$  was  $2\text{ }\mu\text{W m}^{-3}$  at the surface and reduced exponentially with depth, with a drop-off length of 15 km, to 35 km. Between depths of 35 and 70 km,  $A_{\text{rad}}$  was constant at  $0.2\text{ }\mu\text{W m}^{-3}$ , and between 70 and 130 km it was constant at  $0.02\text{ }\mu\text{W m}^{-3}$ . For the mantle, the temperature-dependent values of  $C_p$  and  $\kappa$  were assumed to be those of olivine. Strain heating ( $A_s$ ) occurred within a 3-km-wide shear zone at a depth



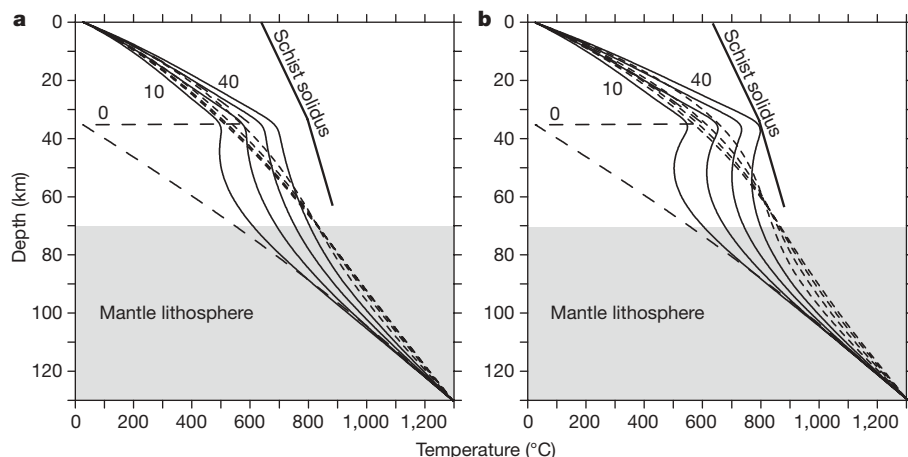
**Figure 2 | Thermal conductivity as a function of temperature.** Curve for average crust calculated using equations (1), (2), (3) and (4) and a density of  $2,700\text{ kg m}^{-3}$ . Curve for olivine calculated from ref. 13. Curves for upper and lower crust calculated from ref. 17; dotted curves include pressure dependence of  $k$ , assuming a geotherm of  $15\text{ }^{\circ}\text{C km}^{-1}$ , and solid curves ignore pressure dependence. Curves for  $\text{KAlSi}_3\text{O}_8$  glass and liquid are from ref. 28. Inset, contrasting temperature dependence of thermal diffusivity and specific heat capacity, calculated for average crust from equations (1), (2), (3) and (4).

of 35 km. The rate of strain heating was given by  $\tau v/z$  where  $\tau$  (shear stress) is 30 MPa,  $v$  (thrusting velocity) is  $3\text{ cm yr}^{-1}$  and  $z$  denotes the characteristic width of the shear zone. The effect of strain heating decreased in a Gaussian fashion with distance from the centre of the shear zone ( $dz$ ) according to  $\exp(-dz^2/z^2)$ . Temperature was fixed at  $1,300\text{ }^{\circ}\text{C}$  at the bottom of 130-km-thick lithosphere. Thrusting was assumed to occur for 40 Myr, followed by 40 Myr of thermal relaxation.

Models with a constant  $\kappa$  of  $1\text{ mm}^2\text{ s}^{-1}$  (Fig. 3a) show that the schist solidus<sup>18,26</sup> is not reached during 40 Myr of thrusting, and thermal relaxation towards a steady-state geotherm after thrusting does not result in attainment of the temperatures required for crustal melting. Incorporating a temperature-dependent  $\kappa$  accelerates the effect of strain heating because heat does not diffuse efficiently away from the hot shear zone (Fig. 3b). In this model, the schist solidus is reached after 40 Myr. Models with constant radiogenic heat production throughout the lithosphere, for example of  $0.1\text{ }\mu\text{W m}^{-3}$  (not shown), produce straighter steady-state geotherms, but the effect of a temperature-dependent  $\kappa$  on strain heating is similar.

Our models indicate that the high degree of thermal insulation provided by hot rocks leads to retention of heat generated by strain heating, and that unusually high radiogenic heat production at mid-crustal levels is not required for melting in orogenic belts. We emphasize that the amounts of radiogenic heating used in our models are conservative, and that higher values will lead to earlier and more widespread crustal melting. Moreover, because thermal diffusion through the hot lithospheric mantle and lower crust is slow, there is a more pronounced curvature to the steady-state geotherm with more potential for melting in the lower crust.

Once melting begins, rock strength decreases and the strain-heating mechanism should cease to be effective<sup>27</sup>. However, the  $\kappa$  of  $\text{KAlSi}_3\text{O}_8$



**Figure 3 | Thermal models for doubly thickened continental crust with a shear zone at a depth of 35 km.** **a**, Model with a constant  $\kappa$  of  $1 \text{ mm}^2 \text{ s}^{-1}$ . Long-dash line is the initial geotherm ( $t = 0$ ), solid lines are geotherms during thrusting at 10-Myr intervals, short-dash lines are relaxation geotherms after cessation of thrusting, also at 10-Myr intervals. Times are indicated in megayears. Schist solidus is from ref. 26. **b**, Model with a temperature-dependent  $\kappa$ , from equations (1) and (2).

glasses and melts is  $\sim 20\%$  lower than that of  $\text{KAlSi}_3\text{O}_8$  crystals<sup>28</sup>, the  $k$  of  $\text{KAlSi}_3\text{O}_8$  glasses and melts is 15–20% lower than that of the bulk crust (Fig. 2), and dissolved water also decreases  $\kappa$  (ref. 29), meaning that hydrous granitic liquids should be particularly efficient insulating materials. The onset of crustal melting may therefore lead to the production of a more insulating layer than the surrounding crust. This positive feedback between melting and thermal insulation may promote increased melt fraction, and allows strain heating to play a significant role in triggering crustal anatexis even though it must become negligible once melting is achieved.

The results of our study pertain to other settings in which partial melting occurs. In both arc and intraplate volcanic settings, basaltic intrusions supply the heat necessary for crustal melting and production of voluminous silicic magmas. Thermal modelling shows that the low thermal diffusivity of hot crust reduces both the long-term flux of mafic magma needed to produce the observed quantities of silicic magma, and the incubation time between the onset of basalt injection and the first eruption of rhyolites<sup>30</sup>. We find values of  $\kappa$  for crustal rocks that are even lower than used in these models, which should result in shorter calculated incubation times. Furthermore, the positive-feedback relations between heating, increased thermal insulation and partial melting will also occur in the mantle, facilitating basaltic magmatism and planetary-scale differentiation.

Received 4 November 2008; accepted 16 January 2009.

- Thompson, W. The age of the Earth as an abode fitted for life. *Science* **9**, 665–674 (1899).
- Turcotte, D. L. & Schubert, G. *Geodynamics* 2nd edn, Ch. 4, 132–194 (Cambridge Univ. Press, 2001).
- Petitjean, S., Rabinowicz, M., Grégoire, M. & Chevrot, S. Differences between Archean and Proterozoic lithospheres: Assessment of the possible major role of thermal conductivity. *Geochim. Geophys. Geosyst.* **7**, doi:10.1029/2005GC001053 (2006).
- Vosteen, H.-D. & Schellschmidt, R. Influence of temperature on thermal conductivity, thermal capacity and thermal diffusivity for different types of rock. *Phys. Chem. Earth* **28**, 499–509 (2003).
- Mottaghy, D., Vosteen, H.-D. & Schellschmidt, R. Temperature dependence of the relationship of thermal diffusivity versus thermal conductivity for crystalline rocks. *Int. J. Earth Sci.* **97**, 435–442 (2008).
- Stein, C. A. & Stein, S. A model for the global variation in oceanic depth and heat flow with lithospheric age. *Nature* **359**, 123–129 (1992).
- Michaut, C., Jaupart, C. & Bell, D. R. Transient geotherms in Archean continental lithosphere: New constraints on thickness and heat production of the subcontinental lithospheric mantle. *J. Geophys. Res.* **112**, doi:10.1029/2006JB004464 (2007).
- Degiovanni, A., Andre, S. & Maillet, D. in *Thermal Conductivity* 22 (ed. Tong, T. W.) 623–633 (Technomic, 1994).
- Hofmann, R., Hahn, O., Raether, F., Mehling, H. & Fricke, J. Determination of thermal diffusivity in diathermic materials by the laser-flash technique. *High Temp. High Press.* **29**, 703–710 (1997).
- Hofmeister, A. M. Thermal diffusivity of garnets to high temperature. *Phys. Chem. Miner.* **33**, 45–62 (2006).

- Hofmeister, A. M. & Pertermann, M. Thermal diffusivity of clinopyroxenes at elevated temperature. *Eur. J. Mineral.* **20**, 537–549 (2008).
- Schumacher, S. & Breuer, D. Influence of a variable thermal conductivity on the thermochemical evolution of Mars. *J. Geophys. Res.* **111**, doi:10.1029/2005JE002429 (2006).
- Hofmeister, A. M., Pertermann, M. & Branlund, J. M. in *Mineral Physics* (ed. Price, G. D.) 543–578 (Elsevier, 2007).
- Branlund, J. M. & Hofmeister, A. M. Factors affecting heat transfer in natural  $\text{SiO}_2$  solids. *Am. Mineral.* **93**, 1620–1629 (2008).
- Pertermann, M. & Hofmeister, A. M. Thermal diffusivity of olivine-group minerals. *Am. Mineral.* **91**, 1747–1760 (2006).
- Robie, R. A. & Hemingway, B. S. Thermodynamic properties of minerals and related substances at 298.15 K and 1 bar ( $10^5$  Pascals) pressure and at higher temperatures. *Bull. US Geol. Surv.* **2131**, 1–461 (1995).
- Chapman, D. S. & Furlong, K. P. in *Continental Lower Crust* (eds Fountain, D. M., Arculus, R. & Kay, R. W.) 179–199 (Elsevier, 1992).
- Nabelek, P. I. & Liu, M. Petrologic and thermal constraints on the origin of leucogranites in collisional orogens. *Trans. R. Soc. Edinb. Earth Sci.* **95**, 73–85 (2004).
- Whittington, A. G. & Treloar, P. J. Crustal anatexis and its relation to the exhumation of collisional orogenic belts, with particular reference to the Himalaya. *Mineral. Mag.* **66**, 53–91 (2002).
- Jamieson, R. A., Beaumont, C., Fullsack, P. & Lee, B. Barrovian regional metamorphism: where's the heat? *Spec. Publ. Geol. Soc. (Lond.)* **138**, 23–51 (1998).
- Huerta, A. D., Royden, L. H. & Hodges, K. V. The effects of accretion, erosion and radiogenic heating on the metamorphic evolution of collisional orogens. *J. Metamorph. Geol.* **17**, 349–366 (1999).
- England, P. & Molnar, P. Cause and effect among thrust and normal faulting, anatectic melting and exhumation in the Himalaya. *Spec. Publ. Geol. Soc. (Lond.)* **74**, 401–411 (1993).
- Harrison, T. M., Grove, M., Lovera, O. M. & Catlos, E. J. A model for the origin of Himalayan anatexis and inverted metamorphism. *J. Geophys. Res.* **103**, 27017–27032 (1998).
- Nabelek, P. I. & Liu, M. Leucogranites in the Black Hills of South Dakota: The consequence of shear heating during continental collision. *Geology* **27**, 523–526 (1999).
- Liu, M. & Furlong, K. P. Crustal thickening and Eocene extension in the southeastern Canadian cordillera: Some thermal and mechanical considerations. *Tectonics* **12**, 776–786 (1993).
- Patiño-Douce, A. E. & Harris, N. Experimental constraints on Himalayan anatexis. *J. Petrol.* **39**, 689–710 (1998).
- Hartz, E. H. & Podladchikov, Y. Y. Toasting the jelly sandwich: The effect of shear heating on lithospheric geotherms and strength. *Geology* **36**, 331–334 (2008).
- Pertermann, M., Whittington, A. G., Hofmeister, A. M., Spera, F. J. & Zayak, J. Transport properties of low-sanidine single-crystals, glasses and melts at high temperature. *Contrib. Mineral. Petrol.* **155**, 689–702 (2008).
- Hofmeister, A. M., Pertermann, M., Branlund, J. M. & Whittington, A. G. Geophysical implications of reduction in thermal conductivity due to hydration. *Geophys. Res. Lett.* **33**, doi:10.1029/2006GL026036 (2006).
- Annen, C., Blundy, J. D. & Sparks, R. S. J. The genesis of intermediate and silicic magmas in deep crustal hot zones. *J. Petrol.* **47**, 505–539 (2006).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank M. Liu for providing the original version of the program OROGEN. This work was supported by the US National Science Foundation.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to A.G.W. ([whittingtona@missouri.edu](mailto:whittingtona@missouri.edu)).

## LETTERS

## Obliquity-paced Pliocene West Antarctic ice sheet oscillations

T. Naish<sup>1,2</sup>, R. Powell<sup>3</sup>, R. Levy<sup>4†</sup>, G. Wilson<sup>5</sup>, R. Scherer<sup>3</sup>, F. Talarico<sup>6</sup>, L. Krissek<sup>7</sup>, F. Niessen<sup>8</sup>, M. Pompilio<sup>9</sup>, T. Wilson<sup>7</sup>, L. Carter<sup>1</sup>, R. DeConto<sup>10</sup>, P. Huybers<sup>11</sup>, R. McKay<sup>1</sup>, D. Pollard<sup>12</sup>, J. Ross<sup>13</sup>, D. Winter<sup>4</sup>, P. Barrett<sup>1</sup>, G. Browne<sup>2</sup>, R. Cody<sup>1,2</sup>, E. Cowan<sup>14</sup>, J. Crampton<sup>2</sup>, G. Dunbar<sup>1</sup>, N. Dunbar<sup>13</sup>, F. Florindo<sup>15</sup>, C. Gebhardt<sup>8</sup>, I. Graham<sup>2</sup>, M. Hannah<sup>1</sup>, D. Hansraj<sup>1,2</sup>, D. Harwood<sup>4</sup>, D. Helling<sup>8</sup>, S. Henrys<sup>2</sup>, L. Hinnov<sup>16</sup>, G. Kuhn<sup>8</sup>, P. Kyle<sup>13</sup>, A. Läufer<sup>17</sup>, P. Maffioli<sup>18</sup>, D. Mages<sup>8</sup>, K. Mandernack<sup>19</sup>, W. McIntosh<sup>13</sup>, C. Millan<sup>7</sup>, R. Morin<sup>20</sup>, C. Ohneiser<sup>5</sup>, T. Paulsen<sup>21</sup>, D. Persico<sup>22</sup>, I. Raine<sup>2</sup>, J. Reed<sup>23,4</sup>, C. Riesselman<sup>24</sup>, L. Sagnotti<sup>15</sup>, D. Schmitt<sup>25</sup>, C. Sjunneskog<sup>26</sup>, P. Strong<sup>2</sup>, M. Taviani<sup>27</sup>, S. Vogel<sup>3</sup>, T. Wilch<sup>28</sup> & T. Williams<sup>29</sup>

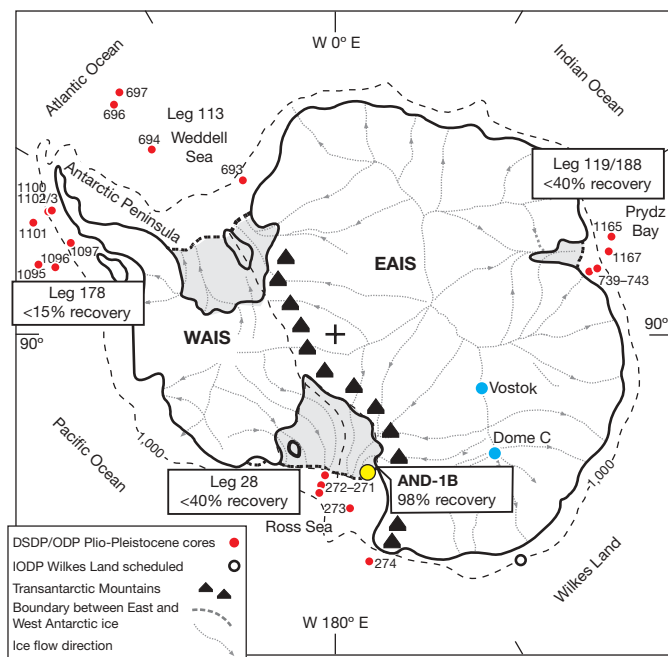
Thirty years after oxygen isotope records from microfossils deposited in ocean sediments confirmed the hypothesis that variations in the Earth's orbital geometry control the ice ages<sup>1</sup>, fundamental questions remain over the response of the Antarctic ice sheets to orbital cycles<sup>2</sup>. Furthermore, an understanding of the behaviour of the marine-based West Antarctic ice sheet (WAIS) during the 'warmer-than-present' early-Pliocene epoch (~5–3 Myr ago) is needed to better constrain the possible range of ice-sheet behaviour in the context of future global warming<sup>3</sup>. Here we present a marine glacial record from the upper 600 m of the AND-1B sediment core recovered from beneath the northwest part of the Ross ice shelf by the ANDRILL programme and demonstrate well-dated, ~40-kyr cyclic variations in ice-sheet extent linked to cycles in insolation influenced by changes in the Earth's axial tilt (obliquity) during the Pliocene. Our data provide direct evidence for orbitally induced oscillations in the WAIS, which periodically collapsed, resulting in a switch from grounded ice, or ice shelves, to open waters in the Ross embayment when planetary temperatures were up to ~3 °C warmer than today<sup>4</sup> and atmospheric CO<sub>2</sub> concentration was as high as ~400 p.p.m.v. (refs 5, 6). The evidence is consistent with a new ice-sheet/ice-shelf model<sup>7</sup> that simulates fluctuations in Antarctic ice volume of up to +7 m in equivalent sea level associated with the loss of the WAIS and up to +3 m in equivalent sea level from the East Antarctic ice sheet, in response to ocean-induced melting paced by obliquity. During interglacial times, diatomaceous sediments indicate high surface-water productivity, minimal summer sea ice and air temperatures above freezing, suggesting an additional influence of surface melt<sup>8</sup> under conditions of elevated CO<sub>2</sub>.

The Earth's climate system during the Pliocene and early-Pleistocene epochs was regulated by a ~40-kyr periodicity. The geological evidence for this is widespread and expressed in polar to equatorial depositional environments including (1) ice volume from oxygen isotope ( $\delta^{18}\text{O}$ ) records, that co-vary with the pattern of ice-rafted debris in deep-sea sediments<sup>9</sup>; (2) ocean circulation<sup>10</sup> and temperature also inferred from deep-sea sediment proxies<sup>11,12</sup>; (3) atmospheric circulation from continental dust deposits<sup>13</sup>; and (4) global sea-level fluctuations recorded in the shallow-marine continental margins<sup>14</sup>.

The 40-kyr cycle is almost certainly linked to variations in the Earth's orbital obliquity. However, the specific nature of this forcing and its influence on Antarctic glaciation remain unresolved owing to a lack of well-dated climate records that directly sample past oscillations of the ice sheet. The new AND-1B core provides such a record (Fig. 1). In this Letter, we focus on the early Pliocene (~5–3 Myr ago) part of the record, because for this period the response of Antarctic ice sheets to orbital forcing can be studied without the complicating influence of large Northern Hemisphere ice sheets on sea-level and deep-sea  $\delta^{18}\text{O}$  records<sup>15</sup>. Furthermore, polar ice-sheet boundary conditions were similar to today, but the climate was warmer<sup>3,5,6</sup>. With anthropogenic warming projected to rise an average of ~3 °C in mean temperature by the end of the twenty-first century, more significance is being placed on the early Pliocene as an analogue for understanding the future behaviour of the WAIS<sup>3</sup> and its contribution to global sea level<sup>16</sup>. Far-field geological evidence for palaeoshorelines up to 25 m above present<sup>17,18</sup> are consistent with ice-volume estimates from deep-ocean  $\delta^{18}\text{O}$  data<sup>19</sup>, and imply deglaciation of the Greenland ice sheet, the

<sup>1</sup>Antarctic Research Centre, Victoria University of Wellington, Kelburn Parade, PO Box 600, Wellington 6012, New Zealand. <sup>2</sup>GNS Science, 1 Fairway Drive, PO Box 30-368, Lower Hutt 5040, New Zealand. <sup>3</sup>Department of Geology & Environmental Geosciences, Northern Illinois University, DeKalb, Illinois 60115, USA. <sup>4</sup>ANDRILL Science Management Office, Department of Geosciences, University of Nebraska-Lincoln, Lincoln, Nebraska 68588-0340, USA. <sup>5</sup>University of Otago, Department of Geology, PO Box 56, Leith Street, Dunedin, Otago 9001, New Zealand. <sup>6</sup>Università di Siena, Dipartimento di Scienze della Terra, Via Laterina 8, I-53100 Siena, Italy. <sup>7</sup>Ohio State University, Department of Geological Sciences, 275 Mendenhall Lab, 125 South Oval Mall, Columbus, Ohio 43210, USA. <sup>8</sup>Alfred Wegener Institute, Department of Geosciences, Postfach 12 01 6, Am Alten Hafen 26, D-27515 Bremerhaven, Germany. <sup>9</sup>Istituto Nazionale di Geofisica e Vulcanologia, Via della Faggiola 32, I-56126 Pisa, Italy. <sup>10</sup>Department of Geosciences, 233 Morrell Science Centre, University of Massachusetts, Amherst, Massachusetts 01003-9297, USA. <sup>11</sup>Department of Earth and Planetary Sciences, Harvard University, Massachusetts 02138, USA. <sup>12</sup>Earth and Environmental Systems Institute, Pennsylvania State University, University Park, Pennsylvania 16802, USA. <sup>13</sup>New Mexico Institute of Mining & Technology, Earth & Environmental Sciences, Socorro, New Mexico 87801, USA. <sup>14</sup>Department of Geology, Appalachian State University, ASU Box 32067, Boone, North Carolina 28608-2067, USA. <sup>15</sup>Istituto Nazionale di Geofisica e Vulcanologia, Via di Vigna Murata 605, I-00143 Rome, Italy. <sup>16</sup>Department of Earth and Planetary Sciences, Johns Hopkins University, Baltimore, Maryland 21218, USA. <sup>17</sup>Federal Institute of Geosciences & Natural Resources, BGR, Stilleweg 2, D-30655 Hannover, Germany. <sup>18</sup>Università Milano-Bicocca, Dipartimento di Scienze Geologiche e Geotecnologie, Piazza della Scienza 4, I-20126 Milano, Italy. <sup>19</sup>Colorado School of Mines, Department of Chemistry & Geochemistry, 1500 Illinois Street, Golden, Colorado 80401, USA. <sup>20</sup>US Geological Survey, Mail Stop 403, Denver Federal Center, Denver, Colorado 80225, USA. <sup>21</sup>University of Wisconsin-Oshkosh, Department of Geology, 800 Algoma Boulevard, Oshkosh, Wisconsin 54901, USA. <sup>22</sup>Dipartimento di Scienze della Terra, Università degli Studi di Parma, Via Usberti 157/A, I-43100 Parma, Italy. <sup>23</sup>CHRONOS, Iowa State University, Department of Geological & Atmospheric Sciences, 275 Science I, Ames, Iowa 50011-3212, USA. <sup>24</sup>Department of Geological and Environmental Sciences, School of Earth Sciences, Stanford University, Stanford, California 94305, USA. <sup>25</sup>Department of Physics, Mailstop #615, University of Alberta, Edmonton, Alberta T6G 2G7, Canada. <sup>26</sup>Department of Geology and Geophysics, Louisiana State University, Baton Rouge, Louisiana 70803, USA. <sup>27</sup>CNR, ISMAR – Bologna, Via Gobetti 101, I-40129 Bologna, Italy. <sup>28</sup>Albion College, Department of Geology, Albion, Michigan 49224, USA. <sup>29</sup>Columbia University, Lamont-Doherty Earth Observatory, Palisades, New York 10964, USA. †Present address: GNS Science, 1 Fairway Drive, PO Box 30368, Lower Hutt 5040, New Zealand.





**Figure 1 | Location of the ANDRILL McMurdo Ice Shelf Project AND-1B drill site in the northwestern corner of the Ross ice shelf.** Also shown are the locations of previous Deep Sea Drilling Project (DSDP), Integrated Ocean Drilling Program (IODP) and Ocean Drilling Program (ODP) cores, Plio-Pleistocene cores, percentage recovery (from shelf sites) and geographic features including ice-sheet configuration and flow lines. (Supplementary Figs 1 and 2 show more detail of the glaciologic and geologic setting). The dashed contour indicates a depth of 1,000 m below sea-level.

WAIS and the marine margins of the East Antarctic ice sheet (EAIS) during the warmest early Pliocene interglacials.

Thirty-eight glaciomarine cycles, each bounded by glacial surfaces of erosion (GSEs), occur in the upper 600 m of the core, and record oscillations in the extent of an ice sheet in Ross embayment during the past 5 Myr (Fig. 2)<sup>20</sup>. The drilled strata accumulated in the rift axis of the Victoria Land basin, ~100 km seaward of the coast. Accommodation space for the preservation of the sediments, and their protection from subsequent glacial erosion, was provided by high rates of tectonic subsidence due to a unique combination of regional rift extension and flexure of the crust by local volcanic islands (Supplementary Information).

Figures 3 and 4 summarize our interpretations of individual sedimentary cycles in terms of the vertical occurrences of lithofacies, that is, sediments representing specific environments of deposition. These range from marine-diatom-rich deposits and mudstones deposited during interglacials to ice-proximal diamictites, conglomerates and sandstones representing glacial periods. During glacial periods, the ice sheet had a laterally extensive marine terminus extending well beyond the drill site, out into the Ross Sea. During interglacials, the drill site was either covered by an ice shelf or, when the ice sheet retreated, lay in open water. The sedimentary characteristics of the cycles and the approach used for their interpretation are given in more detail elsewhere<sup>21</sup> (Methods). We note here that changes in lithofacies through time primarily reflect the proximity of the ice-sheet grounding zone and the thermal characteristics of the depositing ice (Fig. 2 and Supplementary Information). Such inferences are consistent with depositional models from a variety of different glaciomarine regimes<sup>22,23</sup>, and permit the identification of 38 oscillations in the extent of the ice sheet's grounding line.

The composition of till (ice-contact diamictites), which overlie GSEs and represent sediments transported at the base of a grounded ice sheet, show that the ice originated from large outlet glaciers in the Transantarctic Mountains (TAM), especially the Mulock and Skelton

glaciers south of Minna Bluff<sup>21</sup>. A new continental Antarctic ice-sheet model<sup>7</sup> run for the past 5 Myr supports geological interpretations that the provenance of grounded ice at the AND-1B site is always from nearby southern-TAM outlet glaciers during glacial advances (Fig. 2).

Notably, the model finds that local ice variations at the AND-1B site are indicative of the overall West Antarctic glacial state, because both are controlled by variations in ocean-induced melt. When open-water marine sediments occur in the AND-1B core, the model shows not only deglaciation in the western Ross embayment, but also the collapse of the entire WAIS (Fig. 2). Pliocene–Pleistocene variations in ice volume are dominated by large WAIS advances and retreats, while the high-altitude regions of the EAIS remain relatively stable. This is because air temperatures never become warm enough to cause significant surface melting on the EAIS<sup>24</sup>, whereas variations in ocean-induced melt and sea level affect the marine-based WAIS much more than the EAIS. Thus, the sedimentary cycles in the AND-1B core both track local variations of the coastal margin of the EAIS (for example TAM outlet glaciers) and provide physical evidence for major changes in the mass balance of the WAIS (Supplementary Information).

Figure 2 illustrates the stratigraphic position of 26 chronological datums that are used to constrain the age and duration of the 38 sedimentary cycles and identify the time missing at cycle-bounding erosion surfaces, that is, unconformities (Supplementary Table 1). The chronology is developed from <sup>40</sup>Ar/<sup>39</sup>Ar ages of volcanic deposits and a quantitative diatom biostratigraphy used to constrain the correlation between the magnetic polarity stratigraphy and the geomagnetic polarity timescale<sup>21</sup>. The approach used, and associated uncertainties, are outlined in Methods. About 36% of the last 5 Myr is represented as rock in the AND-1B core; the rest is lost at unconformities resulting from erosion through long-term tectonic influences and shorter-term volcanic and glacial processes. Chronostratigraphic constraints enable identification of two types of unconformities: (1) those where the time missing is longer than a Milankovitch cycle, interpreted as major erosion due to tectonic influences and/or a major phase of glacial advance; and (2) those of suborbital duration reflecting lesser glacial erosion associated with a single glacial advance truncating only part of the previous cycle. The chronology also allows the duration of relatively continuous stratigraphic packages comprising more than one cycle to be estimated with sufficient precision for the recognition of orbital periods<sup>25</sup> (Methods).

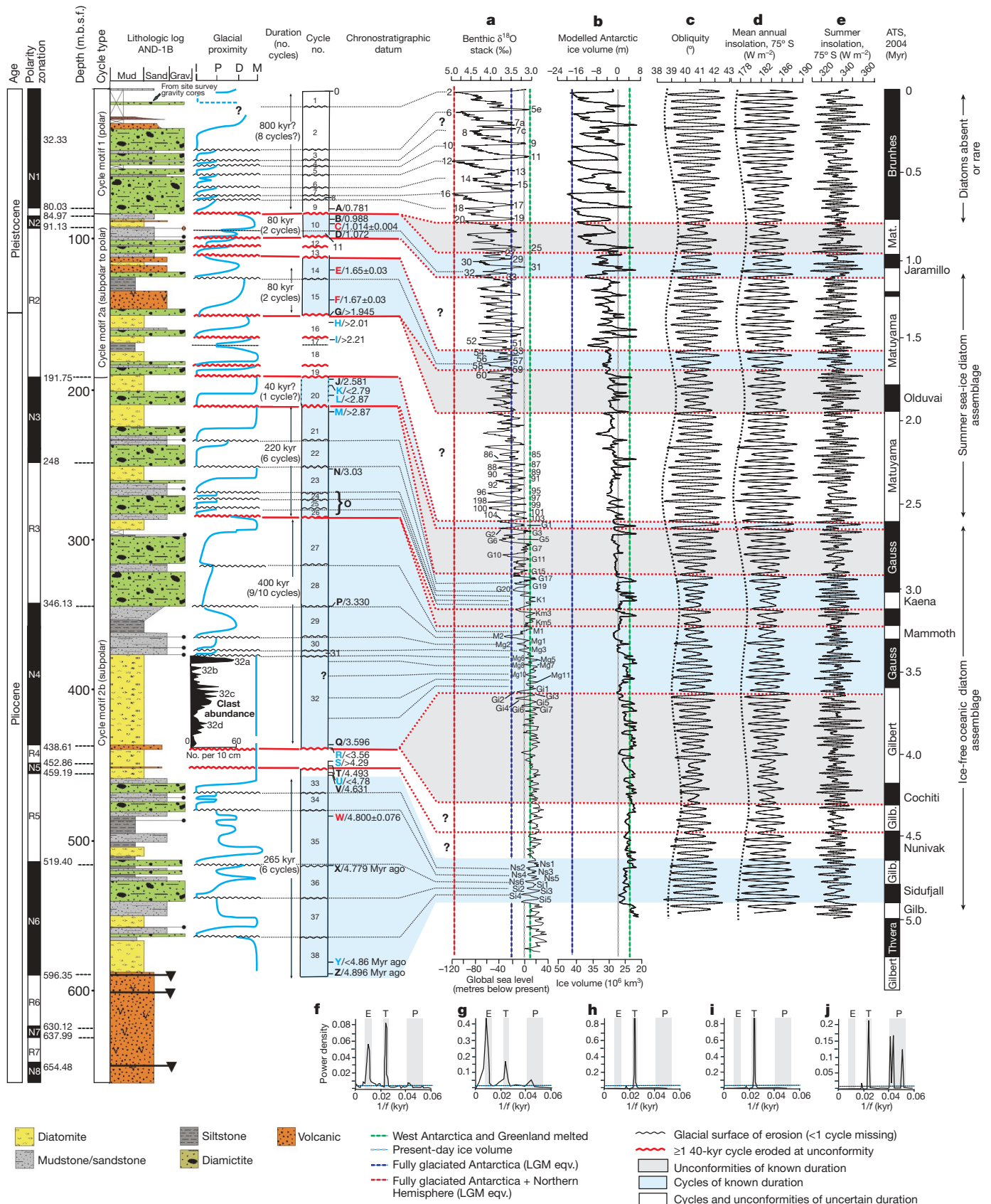
One such interval is illustrated in detail in Fig. 3 and comprises six early-Pliocene glacial cycles (cycles 38–33), spanning 265 kyr. In this case, the identification of three palaeomagnetic reversal boundaries allows one-to-one matching of the WAIS grounding-line oscillations recorded in the AND-1B core to individual 40-kyr ice-volume cycles in the deep-sea benthic  $\delta^{18}\text{O}$  stack<sup>26</sup> and modelled ice-volume cycles between 4.896 and 4.631 Myr ago. Figure 4 illustrates an interval in which the chronology constrains 16 successive early- to mid-Pliocene (3.60–2.87 Myr ago) cycles (32a–d to 20), spanning 700 kyr. In this case, the cyclostratigraphic interpretation and the distribution of time within the AND-1B core are not so straightforward, primarily because a ~60-m-thick interval of marine diatomite, spanning ~200 kyr, occurs between 438 and 376 m.b.s.f. We use the distribution of >2-mm-diameter iceberg-rafted debris (IBRD) through the continuous, thick diatomite unit as an index for ice-rafting intensity and glacial variability. The IBRD record identifies four or five main orbital cycles (32a–d) reflecting glacial fluctuations during an extended period of biopelagic deposition in the Ross embayment, when the ice sheet remained landward of the drill site (Fig. 2). The composition of the IBRD reflects oscillations of local outlet glaciers, which remained near the coast with no significant expansion into the Ross Sea<sup>21</sup>.

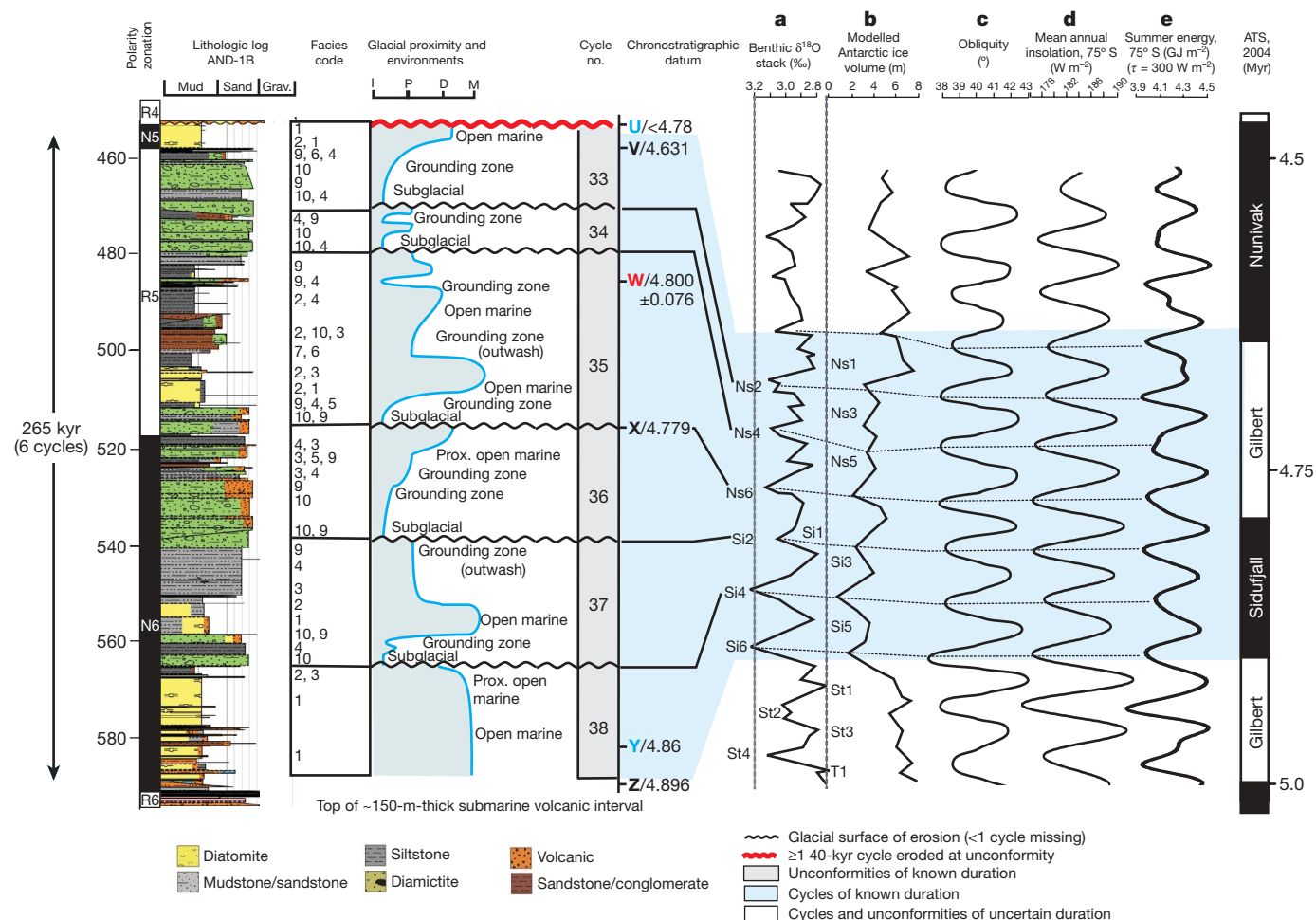
If the four or five IBRD cycles in the diatomite and the eleven overlying unconformity-bounded cycles are distributed evenly over this time interval, the resulting duration is ~40 kyr per cycle (Fig. 4).

The short-duration normal-polarity interval between the Mammoth and Kaena subchrons is not represented in the AND-1B core and could be missing at any of the GSEs between the bases of cycles 28–22. However, if cycles 28–22 are distributed evenly across the amalgamated reversed-polarity subchrons between 3.30 and 3.03 Myr ago, then a

one-to-one match with  $\sim 40$ -kyr  $\delta^{18}\text{O}$  cycles results and suggests that an intervening normal-polarity interval is most likely missing at the basal GSE of cycle 26.

The 60-m-thick diatomite unit lacks a sea-ice-associated diatom flora<sup>27</sup>, and sedimentological evidence implies that warmer-than-present





**Figure 3 | Detailed analysis of early-Pliocene sedimentary cycles in the AND-1B core showing lithofacies interpretations of glacial marine environments.** The glacial proximity curve tracks the relative position of the grounding line through ice-contact (I), ice-proximal (P), ice-distal (D) and marine (M) environments and provides a proxy for ice-sheet extent. Cycle duration is constrained by chronostratigraphic datums coded A–Z. The chronostratigraphy allows a one-to-one correlation of the cycles with

obliquity-paced time series of the benthic  $\delta^{18}\text{O}$  record<sup>26</sup> (a), modelled Antarctic ice volume<sup>7</sup> (expressed in metres of equivalent sea level, b), obliquity (c), mean annual insolation at  $75^\circ\text{S}$  (d), and summer energy at  $75^\circ\text{S}$  for inferred melt threshold ( $\tau$ ) of  $300\text{ W m}^{-2}$  (e; see Supplementary Information for further explanation of summer energy). Dashed vertical lines represent present-day ice volume/sea level. (See Methods and Supplementary Table 2 for explanation of facies codes.)

oceanic and atmospheric conditions existed between  $\sim 3.6$  and  $3.4$  Myr ago. Foraminiferal  $\delta^{18}\text{O}$  values during this time are up to  $0.4\%$  lower than during the Holocene epoch, with amplitudes of  $0.15$  to  $0.35\%$  equivalent to eustatic sea-level fluctuations (including Northern Hemisphere ice; see Methods) of  $+10$ – $30$  m. Modelled Antarctic ice volume reduces to  $20 \times 10^6\text{ km}^3$ , or  $+7$  m in equivalent sea level (Fig. 2), and involves the complete deglaciation of WAIS together with a loss of up to  $+3$  m in equivalent sea level (5%) from the marine margins of the EAIS<sup>7</sup>. If complete deglaciation of Greenland ( $+5$  m) also occurred during this time, then the Antarctic ice-sheet history from isotopes, proximal data and numerical models is in good agreement with the early-Pliocene eustatic sea-level fluctuation amplitudes of  $+10$ – $30$  m

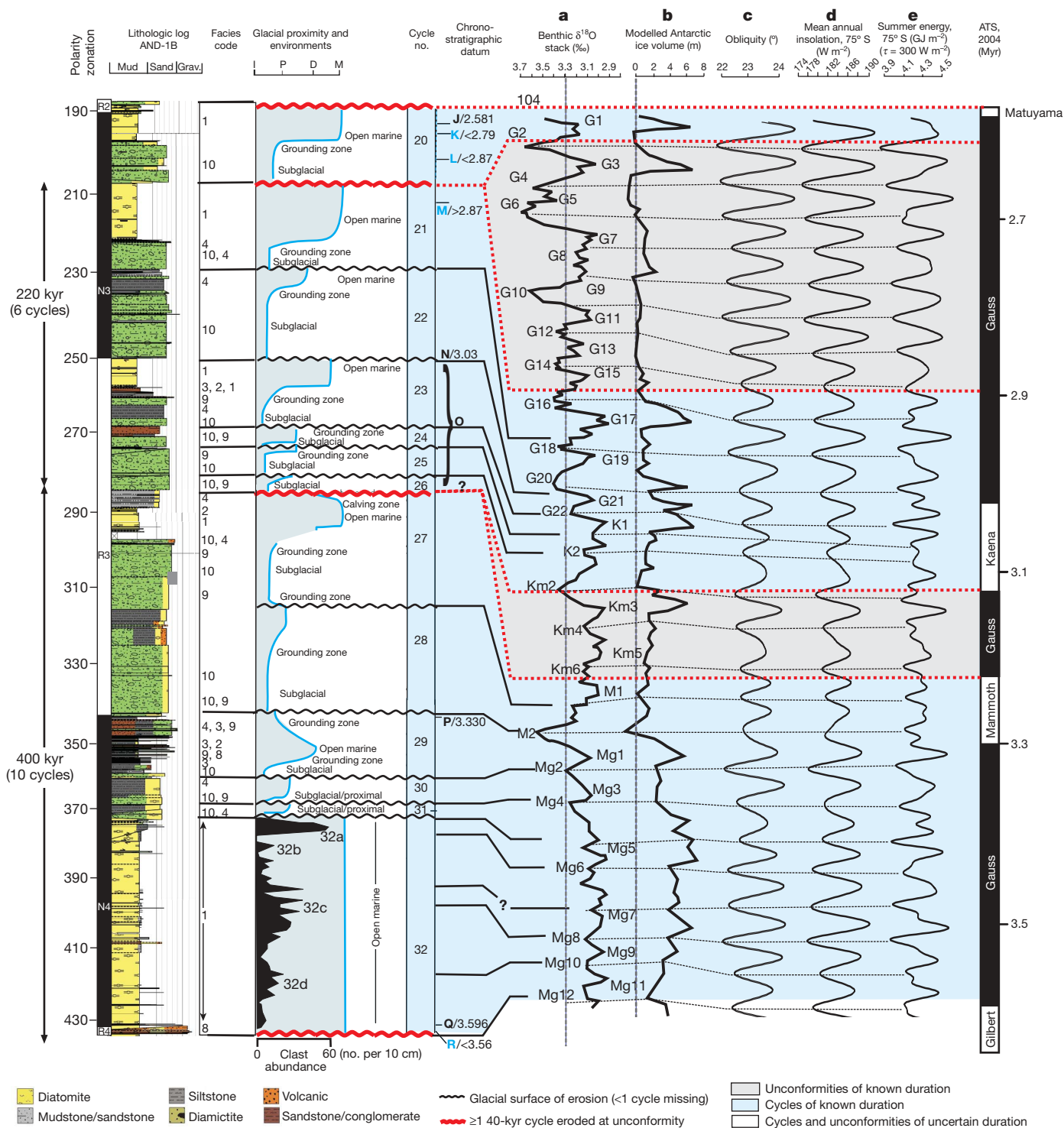
reconstructed from far-field sites such as the uplifted shallow-marine Wanganui basin (New Zealand)<sup>18</sup>.

The WAIS grounding line re-advanced across the western Ross Sea following deposition of the thick diatomite interval, as shown by the occurrence of progressively more ice-proximal facies in the core between sedimentary cycles 31 and 29 (Fig. 4). The interval is correlated with a  $0.7\%$  increase in  $\delta^{18}\text{O}$  between marine isotope stages Mg5 and M2, and a corresponding increase in modelled Antarctic ice volume equivalent to a eustatic fall from  $7$  m above present-day sea level to  $1$  m below<sup>7</sup>. The M2 glacial appears to terminate early-Pliocene warm conditions in the Ross embayment. Our glacial proximity curve shows that WAIS expansion across the drill site occurred during the early

**Figure 2 | Stratigraphic and chronologic summary of the upper 600 m of the AND-1B core showing 38 sedimentary cycles of ice-sheet advance, retreat and re-advance during the last 5 Myr.** Lithologies (rock types) are plotted against depth. The glacial proximity curve tracks the relative position of the grounding line through ice-contact (I), ice-proximal (P), ice-distal (D) and marine (M) environments and provides a proxy for ice-sheet extent. Cycle duration is constrained by chronostratigraphic datums coded A–Z (Supplementary Table 1) and is explained in more detail in Methods. The chronostratigraphy allows correlation of some cycles (blue shading) with time series of the stacked benthic  $\delta^{18}\text{O}$  record<sup>26</sup> ( $\delta^{18}\text{O} = (\delta^{18}\text{O}/^{16}\text{O})_{\text{sample}} / (\delta^{18}\text{O}/^{16}\text{O})_{\text{SMOW}} - 1$ ; SMOW, standard mean ocean water; a), model-output Antarctic ice volume<sup>7</sup> (expressed in metres of equivalent sea level,

b), obliquity (c), mean annual insolation at  $75^\circ\text{S}$  (d) and summer insolation intensity at  $75^\circ\text{S}$  (e). Spectral estimates show that the majority of the variance is at  $\sim 40$ -kyr frequency for the  $\delta^{18}\text{O}$  record (f), obliquity (h) and mean annual insolation (i). Summer insolation at  $75^\circ\text{S}$  has equal power in the obliquity and precession bands (j). The relative increase in 100-kyr power in model-output Antarctic ice volume (g) reflects the increased influence of large Northern Hemisphere ice sheets on ocean temperature and glacioeustatic fluctuations in the late Pleistocene. In f–j: E, eccentricity or  $\sim 100$ -kyr cycle of unknown origin; T, tilt (obliquity); P, precession; f, frequency; vertical axes, power as an arbitrary scale of relative spectral power. m.b.s.f., metres below sea floor; LGM, Last Glacial Maximum. Thin diamictites ( $< 3$  m) are represented by a horizontal line with dot at end.





**Figure 4 | Detailed analysis of late early-Pliocene sedimentary cycles in the AND-1B core showing lithofacies interpretations of glacial marine environments.** Glacial proximity curve tracks the relative position of the grounding line through ice-contact (I), ice-proximal (P), ice-distal (D) and marine (M) environments and provides a proxy for ice-sheet extent. Cycle duration is constrained by chronostratigraphic datums coded A–Z. The

chronostratigraphy allows recognition of 15 cycles, within a ~600-kyr period, that are correlated with obliquity-paced time series of the benthic  $\delta^{18}\text{O}$  record<sup>26</sup> (a), modelled Antarctic ice volume<sup>7</sup> (b), obliquity (c), mean annual insolation at  $75^\circ\text{S}$  (d) and summer energy at  $75^\circ\text{S}$  for  $\tau = 300\text{ W m}^{-2}$  (e). Dashed vertical lines represent present-day ice volume/sea level.

Pliocene when Holocene  $\delta^{18}\text{O}$  values and present-day Antarctic ice volumes ( $27 \times 10^6\text{ km}^3$ ) were exceeded. An up-core transition between 350 and 150 m.b.s.f. is evident, marked by successively fewer submarine outwash deposits and thinner grounding-zone facies successions representing cooling of the ice sheet implied by progressively lower volumes of melt water (Supplementary Information and ref. 21). This culminates with a major cooling step that occurs across the

Gauss–Matuyama polarity transition (2.6 Myr ago) associated with significant erosion and loss of the late-Pliocene stratigraphic record (Fig. 2). This step is taken to represent a major expansion of the ice sheet.

The ~40-kyr-year variability in the size of the WAIS suggests differences from Milankovitch's hypothesis that summer half-year insolation intensity, with its sensitivity to precession (Fig. 2e), should control the growth and decay of ice sheets. Historically, the lack of

precession in the geological record has been attributed to the importance of annual insolation that is controlled by obliquity, with more influence on polar temperatures than seasonal insolation modulated by precession (see, for example, ref. 28). Figure 2 highlights the relationship between obliquity (Fig. 2c), mean annual insolation (Fig. 2d) and the AND-1B cycles for the Pliocene, and Figs 3 and 4 show details for the early Pliocene. Given the sensitivity of WAIS mass balance to ocean temperature<sup>7</sup>, we suggest that 40-kyr orbital cycles may regulate southward export and upwelling of Circumpolar Deep Water with consequences for melt rates at grounding lines of Antarctic ice sheets. Several recent studies have linked changes in Atlantic meridional overturning<sup>29</sup> and Antarctic circumpolar ocean circulation<sup>30</sup> to obliquity forcing. An interglacial mechanism has been proposed whereby the southward expansion of westerly winds and associated northward Ekman transport is compensated for by enhanced upwelling of warmer, CO<sub>2</sub>-rich Circumpolar Deep Water<sup>30</sup>, which also promotes atmospheric warming. Such a positive feedback is supported by the strong correlation between temperature and CO<sub>2</sub> in Antarctic ice-core records and may accelerate sea-ice loss with attendant changes in albedo, further increasing oceanic and atmospheric warming.

The low abundance of sea-ice-associated diatoms (<5%) in the early-Pliocene diatomite intervals of the AND-1B cycles suggests that sea surface and air temperatures may have been above freezing for a significant part of the austral summer. On long timescales, insolation integrated over the length of summer (summer energy) has been shown in models to control the surface melting of ice sheets at the obliquity period<sup>8</sup> (Supplementary Information), provided that the ablating margin is at high latitude and that the surface temperature remains above 0 °C for a significant part of the season<sup>31</sup>. Although the latter condition is not currently met by the Antarctic ice sheet, its surface melt threshold may have been exceeded during the early Pliocene and may be exceeded again in the next 100 years. Given the geological evidence from the AND-1B core, we suggest that the mass balance of the early-Pliocene Antarctic ice sheet, although primarily controlled by ocean-induced melting<sup>7</sup>, may also have been influenced by surface ablation (Figs 3 and 4). A recent model of the early-Pliocene terrestrial EAIS implies sensitivity to surface temperature on its low-elevation margins at an atmospheric CO<sub>2</sub> level of 400 p.p.m. (ref. 32). Furthermore, documented in-phase insolation-linked warming during an early-Pleistocene interglacial<sup>33</sup> suggests that significant melt may occur during precession-amplified obliquity cycles (Supplementary Information). More well-dated sediment cores, and future experiments using ice-sheet models (see, for example, ref. 7), are planned to assess more fully the different influences of orbital forcing on surface ablation versus ocean-induced melt at elevated atmospheric CO<sub>2</sub> levels.

The unconformity-bounded glacial-marine cycles in the AND-1B core provide ice-proximal evidence for ~40-kyr oscillations in Pliocene WAIS and in some of the EAIS outlet glaciers draining into the Ross embayment south of the drill site, a type of behaviour similar to that of the unstable Northern Hemisphere ice sheets of the past ~3 Myr.

## METHODS SUMMARY

The 98%-complete AND-1B drill core was recovered from ~850 m of water, from an 85-m-thick portion of the Ross ice shelf, west of Ross Island (Fig. 1 and Supplementary Fig. 1). Drilling used a custom-built riser system embedded into the sea floor, enabling continuous wireline diamond-bit coring to a depth of 1284.87 m.b.s.f. The core was initially processed at the drill site and transported to McMurdo Station on Ross Island where it was split, logged, sampled and analysed. An initial science report provides a summary of drilling, curation and science methods and the initial results that underpin many of the interpretations in this paper<sup>21</sup>.

The 38 glacial-marine sedimentary cycles are recognized on the basis of the repetitive vertical occurrence of characteristic facies within the AND-1B core. These facies and their interpretations are summarized in Supplementary Table 2, and shown for a single cycle in Supplementary Fig. 3. The age model for the AND-1B

core was developed from quantitative diatom biostratigraphic and <sup>40</sup>Ar/<sup>39</sup>Ar radiometric ages to constrain the correlation between the magnetic polarity stratigraphy and the geomagnetic polarity timescale. Chronostratigraphic data are summarized in Supplementary Table 1. The biostratigraphic and radiometric datums alone are not used for absolute age control or correlation of cycles with the  $\delta^{18}\text{O}$  record, because their numeric ages have uncertainties of up to  $\pm 50$  kyr. However, their precision is sufficient to match polarity zones, identified in the core, with the geomagnetic polarity timescale. These higher-precision ( $\pm 5$  kyr) geomagnetic polarity transitions provide the 'time spikes' for cycle correlation and estimation of cycle duration. For example, the six early-Pliocene glacial cycles between 4.896 and 4.493 Myr ago (Fig. 3) are correlated with C3n.3n/C3n.2r (mid Gilbert, 4.896–4.631 Myr ago) spanning 265 kyr (Fig. 3). Linear interpolation between the dated polarity transitions implies average durations of 39 kyr and 48 kyr per cycle for the lower three (polarity zone N6) and upper three (polarity zone R5) cycles, respectively.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 25 July 2008; accepted 11 February 2009.

- Hays, J. D., Imbrie, J. & Shackleton, N. J. Variations in the Earth's orbit; pacemaker of the ice ages. *Science* **194**, 1121–1132 (1976).
- Raymo, M. E. & Huybers, P. Unlocking the mysteries of the ice ages. *Nature* **415**, 284–285 (2008).
- Intergovernmental Panel on Climate Change in *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Solomon, S. et al.) 707–709 (Cambridge Univ. Press, 2007).
- Kim, S. J. & Crowley, T. J. Increased Pliocene North Atlantic Deep Water: cause or consequence of Pliocene warming. *Paleoceanography* **15**, 451–455 (2000).
- Van Der Burgh, J., Visscher, H., Dilcher, D. & Kürschner, M. Paleotropical signatures in Neogene fossil leaves. *Science* **260**, 1788–1790 (1993).
- Raymo, M. E., Grant, B., Horowitz, M. & Rau, G. H. Mid-Pliocene warmth: stronger greenhouse and stronger conveyor. *Mar. Micropaleontol.* **27**, 313–326 (1996).
- Pollard, D. & DeConto, R. M. Modelling West Antarctic ice sheet growth and collapse through the past five million years. *Nature* (this issue).
- Huybers, P. Early Pleistocene glacial cycles and the integrated summer insolation forcing. *Science* **313**, 508–511 (2006).
- Shackleton, N. J. et al. Oxygen isotope calibration of the onset of ice-rafting and history of glaciation in the North Atlantic region. *Nature* **307**, 620–623 (1984).
- Hall, I. R., McCave, I. N., Shackleton, N. J., Weedon, G. P. & Harris, S. E. Intensified deep Pacific inflow and ventilation in Pleistocene glacial times. *Nature* **412**, 809–812 (2001).
- Crundwell, M., Scott, G., Naish, T. R. & Carter, L. Glacial–interglacial ocean-climate variability spanning the Mid-Pleistocene transition in the temperate Southwest Pacific, ODP site 1123. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **260**, 202–229 (2008).
- Dwyer, G., Baker, P. & Cronin, T. North Atlantic deepwater temperature change during late Pliocene and late Quaternary climatic cycles. *Science* **270**, 1347–1350 (1995).
- Ding, Z. L. et al. Stacked 2.6-Ma grain size record from the Chinese loess based on five sections and correlation with the deep-sea  $\delta^{18}\text{O}$  record. *Paleoceanography* **17**, 5–21 (2002).
- Naish, T. R. Constraints on the amplitude of late Pliocene eustatic sea-level fluctuations: new evidence from the New Zealand shallow-marine sediment record. *Geology* **25**, 1139–1142 (2007).
- Raymo, M. E., Lisiecki, L. & Nisancioglu, K. Plio–Pleistocene ice volume, Antarctic climate, and the global  $\delta^{18}\text{O}$  record. *Science* **313**, 492–495 (2006).
- Mercer, J. H. West Antarctic ice sheet and CO<sub>2</sub> greenhouse effect: a threat of disaster. *Nature* **271**, 321–325 (1978).
- Dowsett, J. & Cronin, T. M. High eustatic sea level during the middle Pliocene: evidence from the southeastern U.S. Atlantic Coastal Plain. *Geology* **18**, 435–438 (1990).
- Naish, T. R. & Wilson, G. Constraints on the amplitude of Mid-Pliocene (3.6–2.4 Ma) eustatic sea-level fluctuations from the New Zealand shallow-marine sediment record. *Phil. Trans. R. Soc. A* **367**, 169–187 (2009).
- Kennett, J. P. & Hodell, D. A. Evidence for relative climatic stability of Antarctica during the Early Pliocene: A marine perspective. *Geogr. Ann.* **75A**, 202–222 (1993).
- Naish, T. R. et al. in *Antarctica: A Keystone in a Changing World* (eds Cooper, A. K. et al.) 71–82 (Proc. 10th Internat. Symp. Antarctic Earth Sci., National Academies Press, 2008).
- Naish, T. R., Powell, R. D. & Levy, R. H. (eds) *Studies from the ANDRILL, McMurdo Ice Shelf Project, Antarctica - Initial Science Report on AND-1B* (Terra Antarctica Vol. 14, 2007).
- Dunbar, G., Naish, T. R., Powell, R. D. & Barrett, P. J. Constraining the amplitude of late Oligocene bathymetric changes in western Ross Sea during orbitally-induced oscillations in the East Antarctic Ice Sheet: (1) Implications for glacial-marine sequence stratigraphic model. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **260**, 50–65 (2008).

23. Powell, R. D. & Cooper, J. M. A sequence stratigraphic model for temperate, glaciated continental shelves. *Spec. Publ. Geol. Soc. (Lond.)* **203**, 215–244 (2003).
24. Lewis, A. R. *et al.* Mid-Miocene cooling and the extinction of tundra in continental Antarctica. *Proc. Natl Acad. Sci. USA* **105**, 10676–10689 (2008).
25. Naish, T. R., Carter, L., Wolff, E., Pollard, D. & Powell, R. D. in *Developments in Earth & Environmental Sciences* Vol. 8 (eds Florindo, F. & Seigert M.) 465–529 (Elsevier, 2009).
26. Lisiecki, L. E. & Raymo, M. E. A. Pliocene-Pleistocene stack of 57 globally distributed benthic  $\delta^{18}\text{O}$  records. *Paleoceanography* **20**, doi:10.1029/2005PA001153 (2005).
27. McKay, R. *et al.* Retreat of the Ross Ice Shelf since the Last Glacial Maximum derived from sediment cores in deep basins surrounding Ross Island. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **260**, 245–261 (2008).
28. Young, M. & Bradley, R. in *Milankovitch and Climate* (eds Berger, A. *et al.*) 707–713 (Riedel, 1984).
29. Lisiecki, L. E., Raymo, M. E. & Curry, W. B. Atlantic overturning responses to Late Pleistocene climate forcings. *Nature* **456**, 85–88 (2008).
30. Toggweiler, J. R., Russell, J. L. & Carson, S. R. Mid-latitude westerlies, atmospheric  $\text{CO}_2$ , and climate change. *Paleoceanography* **21**, doi:10.1029/2005PA001154 (2007).
31. Huybers, P. & Tziperman, E. Integrated summer insolation forcing and 40,000-year glacial cycles: the perspective from an ice-sheet/energy-balance model. *Paleoceanography* **23**, doi:10.1029/2007PA001463 (2008).
32. Hill, D. J., Haywood, A. M., Hindmarsh, R. C. A. & Valdes, P. J. in *Deep Time Perspectives on Climate Change: Marrying the Signals from Computer Models and Biological Proxies* (eds Williams, M. *et al.*) 517–538 (Micropalaeontol. Soc. Spec. Publ., Geological Society of London, 2007).
33. Scherer, R. P. *et al.* Antarctic records of precession paced, insolation-driven warming during the early Pleistocene Marine Isotope Stage 31. *Geophys. Res. Lett.* **35**, doi:10.1029/2007gl032254 (2008).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** The ANDRILL project is a multinational collaboration between the Antarctic programmes of Germany, Italy, New Zealand and the United States. Antarctica New Zealand is the project operator and developed the drilling system in collaboration with A. Pyne. Antarctica New Zealand supported the drilling team at Scott Base; Raytheon Polar Services Corporation supported the science team at McMurdo Station and the Cray Science and Engineering Laboratory. The ANDRILL Science Management Office at the University of Nebraska-Lincoln provided science planning and operational support. The scientific studies are jointly supported by the US National Science Foundation, the New Zealand Foundation for Research Science and Technology and the Royal Society of New Zealand Marsden Fund, the Italian Antarctic Research Programme, the German Research Foundation and the Alfred Wegener Institute for Polar and Marine Research.

**Author Contributions** All authors contributed to acquisition, analysis and interpretation of data presented in this paper. T.N.: overall coordination of writing,

sedimentology, cyclostratigraphic and climatic interpretations; R.P.: integration, glacial facies, glacial process and interpretations of ice-sheet history; R.L.: integration, biochronology and age-model construction; L.K.: core description and sedimentological interpretation; F.N.: core description and physical properties interpretation; M.P.: petrological interpretation; R.S.: integration, diatom biostratigraphic and environmental interpretations; F.T.: clast abundance, composition and provenance interpretations; G.W.: palaeomagnetic stratigraphy and age-model construction; T. Wilson: core description, structural and tectonic constraints; L.C.: sedimentology & palaeo-oceanographic interpretations; R. McKay: sedimentology, glacial facies interpretations and ice-sheet history; J. Ross:  $^{40}\text{Ar}/^{39}\text{Ar}$  geochronology and age-model construction; D.W.: diatom biostratigraphy and environmental interpretations; P.B.: glacial process and interpretations of ice-sheet history; G.B.: glacial marine sequence stratigraphy and facies interpretations; R.C.: biochronology and age-model construction; E.C.: glacial facies, glacial process and interpretations of ice-sheet history; J.C.: biochronology and age-model construction; R.D.: ice-sheet-model data interpretation and integration; G.D.: core description, facies and sedimentological interpretation; N.D.:  $^{40}\text{Ar}/^{39}\text{Ar}$  geochronology and petrological interpretation; F.F.: palaeomagnetic interpretations and age-model construction; C.G.: core description and physical properties interpretation; I.G.: geochronology and age-model construction; M.H.: biostratigraphy and environmental interpretation; D. Harwood: diatom biostratigraphy and biochronology; D. Hansraj: regional seismic stratigraphic context; D. Helling: geochemical interpretation; S.H.: regional stratigraphic framework and tectonic constraints; L.H.: time-series analysis; P.H.: Milankovitch forcing and palaeoclimatic interpretations; G.K.: geochemical interpretation; P.K.: volcanic petrology and volcanological interpretation; A.L.: core description and structural analysis; P.M.: diatom biostratigraphy and environmental interpretations; D.M.: core description and physical properties interpretation; K.M.: core description; W.M.:  $^{40}\text{Ar}/^{39}\text{Ar}$  geochronology and volcanological interpretation; C.M.: core description and structural analysis; R. Morin: borehole description and down-hole geophysics; C.O.: palaeomagnetic stratigraphy and age-model construction; T.P.: core and description and structural geology; D. Persico: calcareous nannofossil biostratigraphy; D. Pollard: ice-sheet-model data interpretation and integration; J. Reed: core description and visualization; C.R.: diatom biostratigraphy and environmental interpretation; I.R.: palynology and environmental interpretation; D.S.: core and borehole description and structural geology; L.S.: palaeomagnetic stratigraphy and age-model construction; C.S.: diatom biostratigraphy and environmental interpretation; P.S.: foram biostratigraphy and environmental interpretation; M.T.: macrofossil biostratigraphy and environmental interpretation; S.V.: subglacial geological interpretation; T. Wilch: core description and interpretation of volcanoclastic sediments; T. Williams: borehole description and down-hole geophysics.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to T.N. ([tim.naish@vuw.ac.nz](mailto:tim.naish@vuw.ac.nz)).



## METHODS

**Facies analysis of sedimentary cycles and the interpretation of glacial proximity.** The individual glacial marine cycles are interpreted to represent ice advance–retreat–re-advance of the grounding line of a laterally continuous ice sheet in Ross embayment, and (from bottom to top) include the following stratigraphic elements.

(1) Erosion and subglacial deposition by an advancing ice sheet producing a GSE that is sharp, planar to sub-horizontal and truncates underlying deposits. Syndepositional soft-sediment deformation structures, intraclasts, clastic dykes and shearing of lithologies occur above and below the GSE. Subglacial features are also indicated by clast orientation and striation within tills.

(2) An interval of grounding-zone deposition comprising conglomerates and poorly sorted stratified sandstones together with stratified diamictite represent a range of basal meltwater conditions observed from subpolar to polar environments<sup>34</sup>, and are associated with glacial marine processes including subglacial meltwater outwash, proglacial debris-flow deposition, and iceberg rainout. These pass upwards into finer-grained terrigenous sediments representing more distal deposition from turbid plumes and grounding-line fans as the grounding line continues to retreat<sup>35,36</sup>. In some cases, ice-shelf environments are indicated by non-fossiliferous, unbioturbated intervals of interstratified sands and silts occurring stratigraphically between ice-proximal and open-marine facies<sup>37</sup>.

(3) Open-marine hemipelagic and pelagic sedimentation with and without iceberg influence. The distal deposits with least glacial influence are diatomites representing biogenic sediment accumulation in an ice-free, highly productive ocean setting.

(4) A proglacial facies succession of progressively more proximal grounding-zone deposits (as in (2)), represents ice-sheet re-advance and eventual glacial overriding at the drill site. This may or may not be preserved below the GSE, depending on the degree of erosion on the GSE.

Several facies commonly associated with proximal glacial marine deposition are noted in AND-1B (Supplementary Table 2), including mudstone with dispersed clasts, conglomerates, sorted sandstone (with a TAM provenance), conglomerates and rhythmically interlaminated couplets of claystone with either siltstone or very fine-grained sandstone (see, for example, Supplementary Fig. 3). The changing abundance of these facies throughout the core, and comparison with modern-day analogues, provides insight into the past extent of subglacial meltwater processes following the concepts established for the Cape Roberts Project core<sup>38</sup>. For example, the rhythmically interlaminated claystone with silt/sandstone facies in AND-1B are consistent with previously described cyclopsam and cyclopel facies from modern temperate to subpolar glacial marine environments in Alaska and the Greenland margin, where they are deposited in quiet-water basins by suspension settling from meltwater plumes<sup>35,36</sup>. These sediments are unknown in modern-day Antarctic glacial marine settings, yet are common in temperate to subpolar settings<sup>39</sup>.

**Chronostratigraphic constraints.** A preliminary age model for the upper 600 m of the drill core constructed from diatom biostratigraphy, and radiometric ages on volcanic material, allows a unique correlation of ~36% of the magnetic polarity stratigraphy with the geomagnetic polarity timescale. Since publication of this age model<sup>21</sup>, the chronology has been improved by further diatom biostratigraphic analysis of the diatomite interval between 460 and 383 m.b.s.f. We now identify a significant hiatus of ~800 kyr at the base of a volcanoclastic gravity-flow deposit at ~440.12 m.b.s.f. Additionally, a new <sup>40</sup>Ar/<sup>39</sup>Ar potassium feldspar age of  $4.800 \pm 0.076$  Myr on a primary volcanic deposit at 481.80 m.b.s.f. (Supplementary Information) confirms the correlation of our magnetic polarity stratigraphy with the geomagnetic polarity timescale for this early-Pliocene interval of the core. All chronostratigraphic datums used to constrain estimates of cycle duration, and correlations with proxy climate and climate-forcing time series in Fig. 2 and Supplementary Fig. 4 for the upper 600 m of the AND-1B core, are summarized in Supplementary Table 1.

**Early-Pliocene chronostratigraphy between 400 and 600 m.b.s.f.** Biostratigraphic event ages include the maximum and minimum ages identified in the CONOP constrained Southern Ocean diatom biochronologic calibration of ref. 40. The ages for new events described below are determined from recent model

runs<sup>41</sup>. The first occurrence of *Rhizosolenia* sp. D is observed at 440.12 m.b.s.f. The maximum age of this event is well constrained in the latest versions of the CONOP models (which incorporate AND-1B data). *Rhizosolenia* sp. D has been documented in at least seven sections from around the Antarctic<sup>40</sup>. The occurrence of this diatom event, therefore, requires the age of the core at and above 440.12 m.b.s.f. to be <3.58 Myr. The palaeomagnetic zone in this portion of the core (440.12–346.13 m.b.s.f.) is normal and correlates C2An.3n (base of the Gauss chron, 3.596–3.300 Myr ago). The absence of *Fragilariopsis interfrigidaria* and *Fragilariopsis barronii* between 450 and 440.12 m.b.s.f. implies an age-range equivalent depth of 4.72–4.29 m.b.s.f. for this interval and constrains the minimum duration of the unconformity below 440.12 m.b.s.f. to ~800 kyr. The first occurrence of *Thalassiosira striata* is observed at 456.00 m.b.s.f., constraining the maximum age of the core at this level to 4.78 Myr and identifying the short normal-polarity interval between 459.19 and 452.86 m.b.s.f. as the Nunivak subchron (C3n.2n, 4.631–4.493 Myr ago). The radiometric age of  $4.8 \pm 0.076$  Myr on the pumice-rich layer at 481.8 m.b.s.f. constrains the superjacent N–R transition at 519.4 m.b.s.f. to the top of the Sidufjall subchron (4.799 Myr ago). The first occurrence of the marine diatom *Thalassiosira complicata* at 583.64 m.b.s.f. is 4.86 Myr ago and constrains correlation of the R–N transition at 596.35 m.b.s.f. with the base of the Sidufjall subchron (4.896 Myr ago).

**Calibration of the benthic foraminiferal  $\delta^{18}\text{O}$  curve to sea level.** The stacked benthic foraminiferal  $\delta^{18}\text{O}$  record for the last 5 Myr (ref. 25) was converted to a sea-level curve by scaling with a calibration of  $0.011\text{‰ m}^{-1}$ , which is consistent with ice-volume  $\delta^{18}\text{O}$  calibrations derived from a number of sources: uplifted Late Quaternary remnant shorelines<sup>42</sup>, equatorial planktonic foraminiferal isotope records with minimal temperature influence<sup>43</sup> and back-stripped estimates of eustatic sea level from continental margin sequences<sup>44</sup>. The calibration removes a consistent 20% temperature contribution from the total amplitude of glacial–interglacial isotopic variation. For example, the LGM-to-Holocene shift in the stacked isotope curve of 1.7‰, when calibrated to the corresponding sea-level shift estimated from Huon Peninsula palaeoshoreline sequence, is 1.2–1.3‰ or  $0.011\text{‰ m}^{-1}$ . The difference is due to glacial–interglacial deep-ocean temperature changes of 2 °C or 0.3–0.4‰.

34. Dowdeswell, J. A., Elverhøi, A. & Spielhagen, R. Glacial marine sedimentary processes and facies on the Polar North Atlantic margins. *Quat. Sci. Rev.* **17**, 243–272 (1998).
35. Ó Cofaigh, C., & Dowdeswell, J. A. Laminated sediments in glacial marine environments: diagnostic criteria for their interpretation. *Quat. Sci. Rev.* **20**, 1411–1436 (2001).
36. Cowan, E. A., Seramur, K. C., Cai, J. & Powell, R. D. Cyclic sedimentation produced by fluctuations in meltwater discharge, tides and marine productivity in an Alaskan fjord. *Sedimentology* **46**, 1109–1126 (1999).
37. Domack, E. W., Jacobson, E. A., Shipp, S. & Anderson, J. B. Late Pleistocene–Holocene retreat of the West Antarctic ice-sheet system in the Ross Sea, Part 2: Sedimentologic and stratigraphic signature. *Geol. Soc. Am. Bull.* **111**, 1517–1536 (1999).
38. Powell, R. D. *et al.* Facies analysis and depositional environments in CRP-3: implications for Oligocene glacial history. *Terra Antarctica* **8**, 207–217 (2001).
39. Powell, R. D. & Domack, E. W. In *Modern and Past Glacial Environments* (ed. Menzies, J.) Ch. 12, 361–389 (Butterworth-Heinemann, 2002).
40. Cody, R., Levy, R., Harwood, D. & Sadler, P. Thinking outside the zone: high-resolution quantitative biochronology for the Antarctic Neogene. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **260**, 92–121 (2008).
41. Cody, R. *et al.* Quantitative biostratigraphic modelling of the AND-1B drillcore. *Glob. Planet. Change* (submitted).
42. Chappell, J. *et al.* Reconciliation of Late Quaternary sea levels derived from coral terraces at Huon Peninsula with deep sea oxygen isotope records. *Earth Planet. Sci. Lett.* **141**, 227–236 (1996).
43. Tiedemann, R., Sarnthein, M. & Shackleton, N. J. Astronomical timescale for the Pliocene Atlantic  $\delta^{18}\text{O}$  and dust flux records of Ocean Drilling Program Site 659. *Palaeoceanography* **9**, 619–638 (1994).
44. Miller, K. G. *et al.* The Phanerozoic record of global sea-level change. *Science* **310**, 1293–1298 (2005).

# Modelling West Antarctic ice sheet growth and collapse through the past five million years

David Pollard<sup>1</sup> & Robert M. DeConto<sup>2</sup>

The West Antarctic ice sheet (WAIS), with ice volume equivalent to  $\sim 5$  m of sea level<sup>1</sup>, has long been considered capable of past and future catastrophic collapse<sup>2–4</sup>. Today, the ice sheet is fringed by vulnerable floating ice shelves that buttress the fast flow of inland ice streams. Grounding lines are several hundred metres below sea level and the bed deepens upstream, raising the prospect of runaway retreat<sup>3,5</sup>. Projections of future WAIS behaviour have been hampered by limited understanding of past variations and their underlying forcing mechanisms<sup>6,7</sup>. Its variation since the Last Glacial Maximum is best known, with grounding lines advancing to the continental-shelf edges around  $\sim 15$  kyr ago before retreating to near-modern locations by  $\sim 3$  kyr ago<sup>8</sup>. Prior collapses during the warmth of the early Pliocene epoch<sup>9</sup> and some Pleistocene interglacials have been suggested indirectly from records of sea level and deep-sea-core isotopes, and by the discovery of open-ocean diatoms in subglacial sediments<sup>10</sup>. Until now<sup>11</sup>, however, little direct evidence of such behaviour has been available. Here we use a combined ice sheet/ice shelf model<sup>12</sup> capable of high-resolution nesting with a new treatment of grounding-line dynamics and ice-shelf buttressing<sup>5</sup> to simulate Antarctic ice sheet variations over the past five million years. Modelled WAIS variations range from full glacial extents with grounding lines near the continental shelf break, intermediate states similar to modern, and brief but dramatic retreats, leaving only small, isolated ice caps on West Antarctic islands. Transitions between glacial, intermediate and collapsed states are relatively rapid, taking one to several thousand years. Our simulation is in good agreement with a new sediment record (ANDRILL AND-1B) recovered from the western Ross Sea<sup>11</sup>, indicating a long-term trend from more frequently collapsed to more glaciated states, dominant 40-kyr cyclicity in the Pliocene, and major retreats at marine isotope stage 31 ( $\sim 1.07$  Myr ago) and other super-interglacials.

Large-scale modelling of the WAIS requires an ice-sheet model that combines the flow regimes of grounded and floating ice efficiently enough to allow simulations of  $\sim 10^5$  yr or more. This is challenging, because the scaled equations for the two regimes are very different, and near the grounding line they interact in a boundary-layer zone that affects the large-scale dynamics<sup>5</sup>. More rigorous higher-order flow models without separate scalings are currently too computationally expensive for long-term continental applications<sup>13</sup>. Our approach simply combines the scaled sheet and shelf equations<sup>12</sup>, while capturing grounding-line effects by imposing a new mass-flux condition<sup>5</sup>. Other standard model components predict variations in ice thickness, ice temperatures, and bedrock elevation below the ice (see Methods).

The multi-million-year timescales considered here are beyond the capability of most climate models to provide the necessary time-continuous forcings required by the ice sheet model. Instead we use techniques similar to those used in previous studies<sup>6,7</sup> and drive

the model with simple parameterizations of surface mass balance, air temperature and specified sea level. A new parameterization of sub-ice-shelf ocean melt based on modern observations<sup>14–16</sup> accounts for changes in the shape of coastlines and distance from the ice edge to open ocean<sup>17</sup> (see Methods).

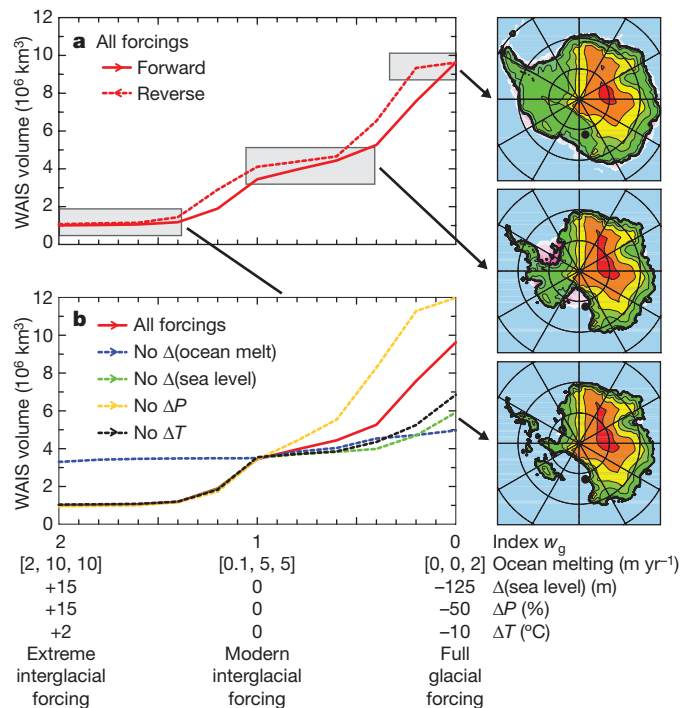
Before considering long-term simulations, it is helpful to examine the link between equilibrated ice-sheet states and the strength of various forcing mechanisms (Fig. 1) representative of extreme interglacial (left of graphs), modern interglacial (middle) and full glacial (right) conditions. In between the values shown, each forcing is linearly interpolated along the  $x$  axis. This closely approximates how they co-vary in long-term simulations, but not exactly due to independent influences of  $\delta^{18}\text{O}$  and austral insolation (see below). The envelopes of ocean-melt values are chosen so that complete WAIS collapse and full glacial expansion are just attained.

Figure 1a indicates a smoothly varying response from intermediate to large WAIS sizes, with sharper transitions into and out of extreme interglacials (collapses), and also back from full glacial to intermediate states. This behaviour is seen in long-term simulations and animations (Supplementary Videos 1, 2), with rapid transitions taking from one to several thousand years. The whole range of Antarctic states in the model is more or less ‘one-dimensional’, that is, the Ross, Weddell and Amundsen Sea sectors of the WAIS usually retreat and expand in unison, resulting in just one type of configuration for a given total ice volume. This suggests that the broad-scale Plio-Pleistocene history of the WAIS is represented at the ANDRILL AND-1B drill site<sup>11</sup>, and persistent absence of a Ross ice shelf is indeed indicative of major WAIS retreat.

The relative importance of individual forcing mechanisms is shown in Fig. 1b. For modern to extreme interglacial conditions, changes in surface climate and sea level are relatively small, while changes in ocean melt are dominant via their effect on ice-shelf buttressing. For modern to glacial conditions, a combination of ocean-melt and sea-level changes is needed to produce realistic WAIS expansion<sup>6,7</sup>. Changes in precipitation and surface temperature have significant, but largely cancelling, effects: without reduced precipitation in cooler climates, glacial volumes are too large (‘no  $\Delta P$ ’, Fig. 1b); without the effects of cooler surface temperatures on internal ice temperatures, viscosities and basal sliding, glacial ice flows too easily and volumes are too small (‘no  $\Delta T$ ’, Fig. 1b).

A five-million-year simulation (Fig. 2) is performed from the early Pliocene to present, with the long-term variation of each forcing mechanism parameterized largely as a function of deep-sea-core  $\delta^{18}\text{O}$  (ref. 18). Sea level over most of this interval is dominated by Northern Hemispheric ice volume, and can be readily prescribed in proportion to  $\delta^{18}\text{O}$ . The responses of Antarctic surface temperature and precipitation to Pleistocene glacial cycles are also reasonably constrained by climate studies and observations, and we adapt established

<sup>1</sup>Earth and Environmental Systems Institute, Pennsylvania State University, University Park, Pennsylvania 16802, USA. <sup>2</sup>Department of Geosciences, University of Massachusetts, Amherst, Massachusetts 01003, USA.

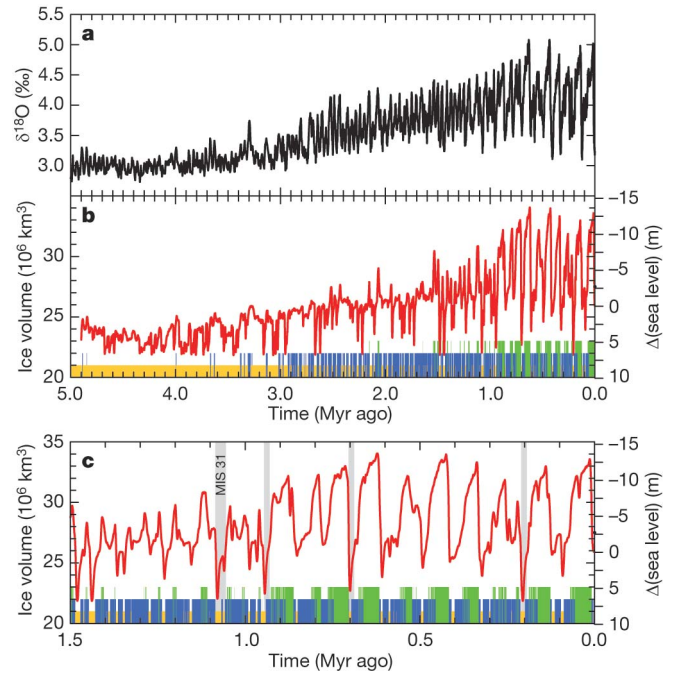


**Figure 1 | Equilibrium West Antarctic ice volumes versus specified forcing, and ice-sheet configurations.** Left panels, ice volumes. The four forcing mechanisms are sub-ice-shelf oceanic melting and departures of sea level, annual precipitation  $\Delta P$  and temperature  $\Delta T$  from present. The three sets of forcing values represent climates for extreme interglacial (left), modern interglacial (middle), and full glacial (right). In between, each forcing is linearly interpolated along the x-axis ( $w_g$ , see equation (6) in Methods). The triplets of sub-ice oceanic melt rates are for protected, exposed-shelf, and deep-ocean regions— $[M_p, M_e, M_d]$ , equations (7) and (8) in Methods. **a**, All forcing mechanisms changed together. Solid (dashed) curves are generated with ice sheets initialized from prior solutions representing cooling (warming) trends. The slight difference between the two curves contrasts with the much larger hysteresis of East Antarctica, where surface melt and not sub-ice-shelf melt is the dominant ablation process<sup>32</sup>. **b**, With one mechanism held constant at its modern value, and all others changed. Right panels, ice sheet configurations representative of the three climatic states, with the black dot showing the location of the ANDRILL AND-1B drill site<sup>11</sup>.

parameterizations using  $\delta^{18}O$  and austral insolation as inputs (equations (1) and (2) in Methods).

Factors controlling past variations of oceanic sub-ice melt on  $\sim 10^4$  yr timescales are less certain. Sub-ice oceanic melting is affected in part by circum-Antarctic deep-water (CDW) warmth and its incursions onto continental shelves<sup>19</sup>. We argue that CDW and sub-ice melt have been mainly controlled by far-field climatic influences that vary in step with Northern Hemispheric glacial–interglacial cycles (see Methods). Without identifying the explicit link (which may involve atmospheric  $CO_2$ , meridional overturning circulation, sea level, or other global-scale teleconnections), we hypothesize that temporal variations of Antarctic sub-ice ocean melt rates are represented by records that correlate with Northern Hemispheric glacial variations, that is, deep-sea-core  $\delta^{18}O$  (equations (6)–(8) in Methods). A minor additional influence on sub-ice melt from austral summer orbital insolation anomalies<sup>20</sup> is also needed to produce precessional cyclicity like that observed during marine isotope stage 31 (MIS 31) around 1 Myr ago<sup>11,21</sup>. Our forcing is warmest during the early Pliocene warm period ( $\sim 5$  to  $\sim 3$  Myr ago) due to light  $\delta^{18}O$  values at that time; however, the parameterizations are based more on Pleistocene variations, and may not fully represent the warm Pliocene if unique processes (for example, persistent El Niño)<sup>9</sup> were involved.

With long-term forcing variations mainly following deep-sea-core  $\delta^{18}O$  (ref. 18), the ice-sheet model is continuously integrated over the past 5 Myr (Fig. 2). Except for small variations along the Wilkes



**Figure 2 | Simulated total Antarctic ice volume over the past five million years.** **a**, Stacked deep-sea-core benthic  $\delta^{18}O$  (ref. 18). **b**, Total Antarctic ice volume (red line) in a long-term simulation with variations of sub-ice melt and other forcings parameterized mainly from the deep-sea-core  $\delta^{18}O$  record. Equivalent changes in global sea level are shown on the right, accounting for the fraction of grounded ice above sea level compared to that below sea level<sup>1</sup>. Bars along the x-axis indicate conditions at a single location ( $78.0^{\circ}S$ ,  $169.4^{\circ}E$ ), shifted one grid box to the east of AND-1B<sup>11</sup> to avoid poorly resolved Ross Island shorelines (yellow, open ocean; blue, floating ice shelf; green, grounded ice). Yellow and blue/green here correspond to the AND-1B diatomite (yellow) and diamicite (green) intervals in Fig. 2 of ref. 11. **c**, As **b** but with the time axis expanded over the past 1.5 Myr. Grey shading indicates simulated super-interglacials, beginning with MIS 31<sup>21</sup>.

margin<sup>22</sup> and in inlets such as Prydz bay<sup>23</sup>, East Antarctica is stable throughout the simulation and nearly all of the ice-volume variability is due to West Antarctica. Several key aspects of the model time series agree with the AND-1B core<sup>11</sup>. There is an overall progression from predominantly smaller WAIS sizes to larger. Furthermore, intervals of WAIS collapse with little or no marine ice are much more common from  $\sim 5$  to 3 Myr ago, which is consistent with intervals in the drillcore dominated by diatomaceous sediments indicating warmer sea surface temperatures, little or no summer sea ice, and an open marine Ross embayment<sup>11</sup>. In fact, the two thickest diatomaceous intervals in the core, between  $\sim 4.3$  and 3.4 Myr ago, correspond to the period with the most frequent and prolonged WAIS collapses simulated by the model. These collapses could well be continuous if additional Pliocene warm-period forcing was added<sup>9</sup>. After 3 Myr ago, there are longer intervals with modern-to-glacial ice volumes, that is, with ice-shelf or grounded-ice cover at or near the AND-1B site (Fig. 2), again in rough agreement with the increasing predominance of diamicite after 3 Myr ago indicating overriding ice or a proximal grounding zone<sup>11</sup>.

Brief WAIS super-interglacial collapses occur after 3 Myr ago but with decreased frequency. In some cases, these precisely match the thinner diatomaceous intervals in the AND-1B core, including the well-dated MIS 31 event at 1.07 Myr ago<sup>11,21</sup>. The large 100-kyr fluctuations of the past million years are similar to those modelled in earlier studies<sup>6,7,17</sup>. The last retreat of WAIS from  $\sim 15$  kyr ago to the present roughly matches the observed retreat of Ross Sea grounding lines<sup>24,25</sup>, and is particularly realistic with modifications described in Supplementary Information section 6.

The model predicts several major WAIS collapses during Pleistocene interglacials (Fig. 2c), at times when  $\delta^{18}O$  minima coincide with strong

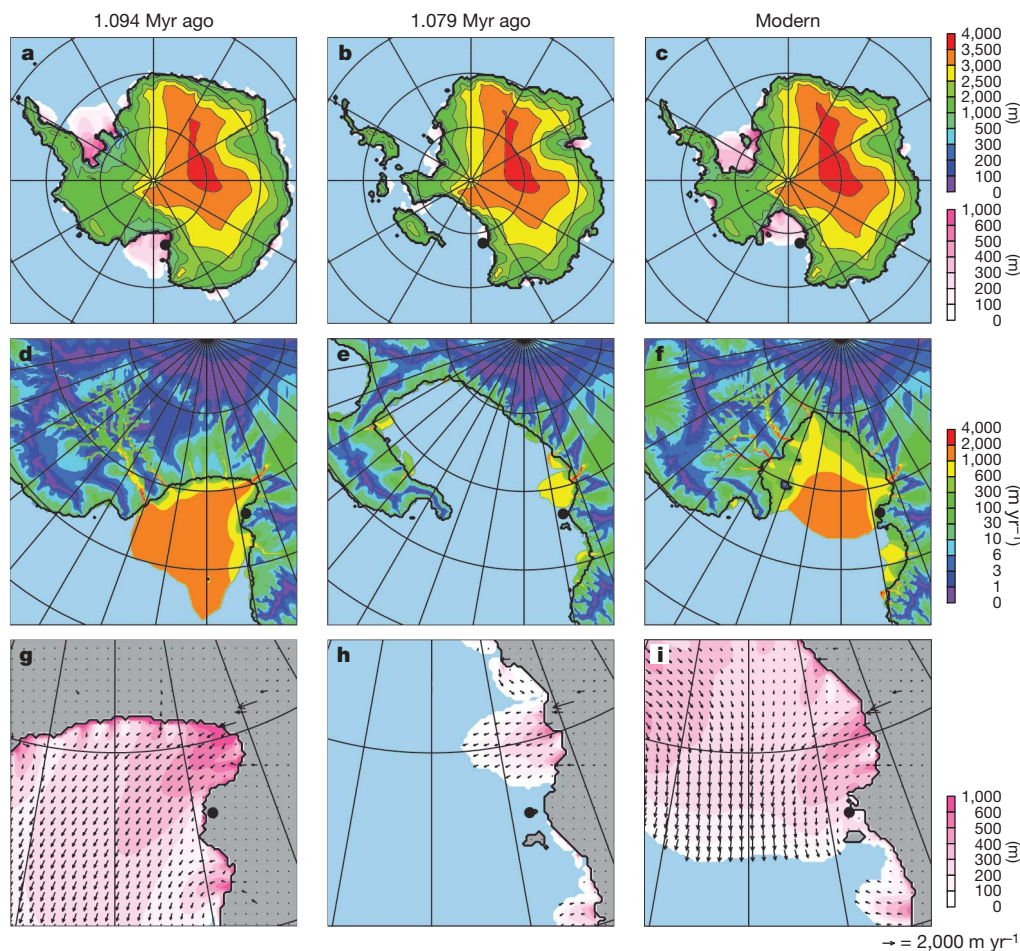


austral summer insolation anomalies. The simulated collapse at MIS 31 corresponds well with core evidence<sup>11,21</sup>, both in terms of timing and magnitude, but more recent collapses (for example, ~200 kyr ago) do not always coincide with the late-Pleistocene interglacials (~125 kyr ago and ~400 kyr ago) usually suspected of harbouring such events<sup>10,26,27</sup>. Thus, while the total number of collapses is reasonable, their sometimes imprecise timings may reflect the limitations of our simple forcing parameterizations, including uncertainties in the 40-kyr phase relationship of Antarctic sub-ice melt to deep-sea-core  $\delta^{18}\text{O}$  records, and the influence of local orbital insolation forcing. Although the model clearly captures the overall ~40-kyr periodicity seen in the AND-1B record, the precise phasing between Antarctic ice sheet variations and Northern Hemispheric climate changes remains uncertain. In some instances, the timing of our simulated super-interglacials may be an artefact of the phasing between the imposed  $\delta^{18}\text{O}$  and austral summer insolation forcings. Recent observational and modelling studies on the relative timing of Northern Hemisphere ice volume variations, ocean meridional overturning and orbital forcing<sup>28–30</sup> are pertinent to this issue, but with no clear consensus to date. These relationships could also be explored in future work with global climate models in combination with regional circum-Antarctic and sub-ice-shelf ocean modelling<sup>14,19</sup>, to better ascertain the effects of Northern Hemispheric glacial cycles, orbital forcing and greenhouse gas concentrations on regional Antarctic conditions.

To better focus on the Ross embayment and the AND-1B site<sup>11</sup>, we ran higher-resolution (10 km) nested ice sheet-shelf simulations for particular times, with boundary conditions at the domain edges

obtained from the long-term all-Antarctic simulation. Figure 3 illustrates a wide range of WAIS states, from weak glacial, full WAIS collapse, to modern conditions. The modern network and behaviour of Siple Coast ice streams and Transantarctic outlet glaciers is well resolved (Fig. 3f), with some ice streams stagnating and re-activating over the several thousand years of the nested run<sup>31</sup> (Supplementary Videos 3 and 4). Ross ice shelf velocities are also similar to observations, as is the central streamline dividing Siple (West Antarctic) and Transantarctic (East Antarctic) ice<sup>31</sup>. The finer ice grid resolves the general ice flow around Ross Island, although the details of flow are not fully resolved within the narrow confines of McMurdo Sound containing AND-1B and other drill sites<sup>11,21</sup>. When shelf ice is present at AND-1B (Fig. 3i), offshore flow just to the east is always northward, with ice originating from major Transantarctic outlet glaciers to the south (Byrd, Skelton, Mulock). This offshore flow pattern and its Transantarctic provenance prevail whenever there is shelf ice around Ross Island.

The dominant regional control is the overall strength of sub-ice oceanic melting in the Ross embayment, which causes both Siple- and Transantarctic-sourced ice to recede or advance in concert over the eastern and western sides of the embayment, respectively. It is very rare for one type or the other to dominate. Thus, although the provenance of shelf ice around Ross Island may be insensitive to the overall WAIS state, the basic presence or absence of shelf ice at the AND-1B site<sup>11</sup> is a good qualitative indicator of maxima and minima in WAIS ice volume (Fig. 2). Other sites not yet cored in the central Ross embayment may offer even better potential for uniquely identifying times of WAIS collapse (Supplementary Fig. 3). These simulations show how



**Figure 3 | Snapshots at particular times from the long-term simulation in Fig. 2.** Shown are 1.094 Myr ago, 1.079 Myr ago (MIS 31 retreat) and modern. **a–c**, Grounded ice elevations and floating ice thicknesses, shown respectively (in m) by upper and lower colour scale on right. **d–f**, Surface ice speeds ( $\text{m yr}^{-1}$ ), from higher-resolution (10 km) nested runs over the Ross

embayment for the same three times, showing the whole nested domain. **g–i**, Floating ice thicknesses (m) and velocity vectors from the nested simulations, enlarged over the western Ross embayment. Vectors are shown only every third grid point for clarity. The location of AND-1B is shown by the black dot.

local observables in the AND-1B and other cores relate to overall WAIS evolution. In particular, our results imply that the presence or absence of grounded or floating ice in the vicinity of McMurdo Sound is indeed linked to WAIS ice volume, and that open-water conditions in the Ross Sea are indicative of partial to complete collapse of the WAIS.

Some of our results are independent of the parameterized temporal variations in long-term forcing. For example, the estimated magnitudes of sub-ice oceanic melt rates needed to produce full WAIS amplitudes (Fig. 1 and Methods) form a point of reference for future modelling. Another independent result is the tendency for the WAIS to experience relatively rapid transitions within one to a few thousand years, as forcing is smoothly varied. This includes transitions into and out of collapsed states, and from full glacial to modern-like ice extents. A collapse from modern conditions occurs when sub-ice ocean melting increases from  $0.1$  to  $2 \text{ m yr}^{-1}$  under shelf interiors, and from  $5$  to  $10 \text{ m yr}^{-1}$  near exposed shelf edges ( $M_p$  and  $M_e$  respectively, in equations (3), (7) and (8) in Methods). Recent melt rates under small Antarctic ice shelves are inferred to be increasing dramatically<sup>15,16</sup>. The relationship between sub-ice melt rates and ocean temperatures is just beginning to be explored<sup>19</sup>, but those data<sup>15,16</sup> and simplified modelling<sup>14</sup> suggest relationships on the order of  $10 \text{ m yr}^{-1} \text{ } ^\circ\text{C}^{-1}$  for smaller shelves, and  $0.4 \text{ m yr}^{-1} \text{ } ^\circ\text{C}^{-1}$  for whole-shelf averages under the major Ross and Filchner-Ronne shelves. Dividing our interior-melt ( $M_p$ ) increase of  $1.9 \text{ m yr}^{-1}$  by the latter sensitivity of  $0.4 \text{ m yr}^{-1} \text{ } ^\circ\text{C}^{-1}$  suggests that the WAIS will begin to collapse when nearby ocean temperatures warm by roughly  $5 \text{ } ^\circ\text{C}$ . Global climate and regional ocean modelling is needed to predict when and if future ocean temperatures and melt rates under the major Antarctic ice shelves will increase by these amounts, and if so, for how long.

## METHODS SUMMARY

The scaled dynamical equations for sheet flow (shallow ice approximation) and shelf flow can be combined heuristically<sup>12</sup>. However, for efficiency in these long-term simulations, they are applied separately depending on whether ice is grounded or floating. Despite this simplification and coarse grids, the effects of the grounding-line boundary layer are captured by imposing a mass-flux condition across the grounding line following ref. 5, which sets ice velocities there as a function of ice thickness. To include important effects of ice-shelf buttressing, the imposed grounding-line velocities are reduced depending on the ratio of longitudinal stress to its free-floating value<sup>5</sup> (see Supplementary Information). The model also contains three other standard components: (1) an ice-mass advection equation predicting ice thickness and accounting for surface accumulation minus ablation and basal melt, (2) an ice temperature equation including horizontal advection, vertical diffusion and shear heating, and (3) a bedrock elevation equation with local relaxation towards isostatic equilibrium and elastic lithospheric flexure<sup>6,7</sup>. There is no explicit basal hydrology, other than allowing basal sliding only where the bed is at the melt point.

Equilibrium ice-free topography and bathymetry are prescribed from the modern BEDMAP database<sup>1</sup>, by removing all ice and allowing the bed to rebound isostatically. Prescribed basal sliding coefficients crudely represent the likely spatial distribution of deformable sediment versus hard bedrock, that is, sediment where the ice-free rebounded topography is below sea level (mostly WAIS) and bedrock where above (mostly East Antarctic ice sheet). In addition, intermediate basal stiffness is prescribed in the Pine Island/Thwaites drainage sector and Transantarctic inlets below sea level, to improve modern grounding-line locations and glacier velocities there. Past surface mass balance and sub-ice-shelf oceanic melting are parameterized using deep-sea-core  $\delta^{18}\text{O}$  and orbital insolation variations (see Methods). The model is run on a polar stereographic grid, with  $40 \text{ km}$  resolution for continental and  $10 \text{ km}$  for nested experiments.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 12 August 2008; accepted 8 January 2009.

1. Lythe, M. B. *et al.* BEDMAP: A new ice thickness and subglacial topographic model of Antarctica. *J. Geophys. Res. Solid Earth* **106** (B6) 11335–11351 (2001).
2. Mercer, J. H. West Antarctic ice sheet and  $\text{CO}_2$  greenhouse effect: A threat of disaster. *Nature* **271**, 321–325 (1978).
3. Weertman, J. Stability of the junction of an ice sheet and an ice shelf. *J. Glaciol.* **13**, 3–11 (1974).

4. Oppenheimer, M. & Alley, R. B. The West Antarctic ice sheet and long term climate policy – An editorial comment. *Clim. Change* **64**, 1–10 (2004).
5. Schoof, C. Ice sheet grounding line dynamics: Steady states, stability, and hysteresis. *J. Geophys. Res.* **112**, F03S28, doi:10.1029/2006JF000664 (2007).
6. Ritz, C., Rommelaere, V. & Dumas, C. Modeling the evolution of Antarctic ice sheet over the last 420,000 years: Implications for altitude changes in the Vostok region. *J. Geophys. Res.* **106** (D23) 31943–31964 (2001).
7. Huybrechts, P. Sea-level changes at the LGM from ice-dynamic reconstructions of the Greenland and Antarctic ice sheets during the glacial cycles. *Quat. Sci. Rev.* **21**, 203–231 (2002).
8. Anderson, J. B. *Antarctic Marine Geology* (Cambridge Univ. Press, 1999).
9. Ravelo, A. C. *et al.* Regional climate shifts caused by gradual global cooling in the Pliocene epoch. *Nature* **429**, 263–267 (2004).
10. Scherer, R. P. Quaternary and Tertiary microfossils from beneath Ice Stream-B – Evidence for a dynamic West Antarctic Ice-Sheet History. *Glob. Planet. Change* **90**, 395–412 (1991).
11. Naish, T. *et al.* Obliquity-paced, Pliocene West Antarctic ice sheet oscillations. *Nature* (this issue).
12. Pollard, D. & DeConto, R. M. in *Glacial Sedimentary Processes and Products* (eds Hambrey, M., Christoffersen, P., Glasser, N. & Hubbard, B.) 37–52 (Spec. Publ. 39, International Association of Sedimentologists, Blackwell Publishing, 2007).
13. Pattyn, F. A new three-dimensional higher-order thermomechanical ice sheet model: Basic sensitivity, ice stream development, and ice flow across subglacial lakes. *J. Geophys. Res.* **108** (B8), 2382, doi:10.1029/2002JB002329 (2003).
14. Beckmann, A. & Goose, H. A parameterization of ice shelf-ocean interaction for climate models. *Ocean Model.* **5**, 157–170 (2003).
15. Rignot, E. & Jacobs, S. S. Rapid bottom melting widespread near Antarctic ice sheet grounding lines. *Science* **296**, 2020–2023 (2002).
16. Shepherd, A., Wingham, D. & Rignot, E. Warm ocean is eroding West Antarctic Ice Sheet. *Geophys. Res. Lett.* **31**, L23402, doi:10.1029/2004GL021106 (2004).
17. Philippson, G. *et al.* Evolution of the Antarctic ice sheet throughout the last deglaciation: A study with a new coupled climate – north and south hemisphere ice sheet model. *Earth Planet. Sci. Lett.* **248**, 750–758 (2006).
18. Lisiecki, L. E. & Raymo, M. E. A Pliocene-Pleistocene stack of 57 globally distributed benthic  $\delta^{18}\text{O}$  records. *Paleoceanography* **20**, PA1003, doi:10.1029/2005PA001153 (2005).
19. Holland, P. R. & Jenkins, A. The response of ice shelf basal melting to variations in ocean temperature. *J. Clim.* **21**, 2558–2572 (2008).
20. Laskar, J. *et al.* A long-term numerical solution for the insolation quantities of the Earth. *Astron. Astrophys.* **428**, 261–285 (2004).
21. Scherer, R. P. *et al.* Antarctic records of precession-paced insolation-driven warming during early Pleistocene Marine Isotope Stage 31. *Geophys. Res. Lett.* **35**, L03505, doi:10.1029/2007GL032254 (2008).
22. Hill, D. J., Haywood, A. M., Hindmarsh, R. C. A. & Valdes, P. J. in *Deep Time Perspectives on Climate Change: Marrying the Signals from Computer Models and Biological Proxies* (eds Williams, M., Haywood, A. M., Gregory, F. J. & Schmidt, D. N.) 517–538 (Micropaleontological Society Special Publications, Geological Society, 2007).
23. O'Brien, P. E. *et al.* Late Neogene ice drainage changes in Prydz Bay, East Antarctica and the interaction of Antarctic ice sheet evolution and climate. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **245**, 390–410 (2007).
24. Conway, H. *et al.* Past and future grounding-line retreat of the West Antarctic Ice Sheet. *Science* **286**, 280–283 (1999).
25. McKay, R. M. *et al.* Retreat history of the Ross Ice Sheet (Shelf) since the Last Glacial Maximum from deep-basin sediment cores around Ross Island. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **260**, 245–261 (2008).
26. Hearty, P. J. *et al.* Global sea-level fluctuations during the Last Interglaciation (MIS 5e). *Quat. Sci. Rev.* **26**, 2090–2112 (2007).
27. Raynaud, D. *et al.* in *Earth's Climate and Orbital Eccentricity: The Marine Isotope Stage 11 Question* (eds Droxler, A. W., Poore, R. Z. & Burckle, L. H.) 27–40 (American Geophysical Union, 2003).
28. Lisiecki, L. E., Raymo, M. E. & Curry, W. B. Atlantic overturning responses to Late Pleistocene climate forcings. *Nature* **456**, 85–88 (2008).
29. Huybers, P. & Denton, G. Antarctic temperature at orbital timescales controlled by local summer duration. *Nature Geosci.* **1**, 787–792 (2008).
30. Kawamura, K. *et al.* Northern Hemisphere forcing of climatic cycles in Antarctica over the past 360,000 years. *Nature* **448**, 912–916 (2007).
31. Hulbe, C. & Fahnestock, M. Century-scale discharge stagnation and reactivation of the Ross ice streams, West Antarctica. *J. Geophys. Res.* **112**, F03S27, doi:10.1029/2006JF000603 (2007).
32. Pollard, D. & DeConto, R. M. Hysteresis in Cenozoic Antarctic ice sheet variations. *Glob. Planet. Change* **45**, 9–21 (2005).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank T. Naish and R. Powell for discussions on this work, and P. Barrett for comments on the manuscript. This work was funded by the US National Science Foundation under awards ATM-0513402/0513421, ANT-034248 and ANT-0424589.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to D.P. (pollard@essc.psu.edu).



## METHODS

**Modern climatic forcing: temperature and precipitation.** Modern forcing fields of annual surface mass-balance and temperature are specified using simple empirical parameterizations, and then varied in the past depending on ice-core or deep-sea-core time series, similarly to previous studies<sup>6,7</sup>. Annual surface temperatures (°C) are<sup>33</sup>

$$T = T_m + 34.46 - 0.00914 h_s - 0.68775 |\phi| + 0.1 \Delta q_a + 10 \Delta s / 125 \quad (1)$$

where  $T_m = 0^\circ\text{C}$ ,  $h_s$  is elevation (m),  $|\phi|$  is latitude (°S),  $\Delta q_a$  is annual orbital insolation anomaly from present at  $80^\circ\text{S}$  ( $\text{W m}^{-2}$ ), and  $\Delta s$  is sea-level departure from present (m) representing atmospheric  $\text{CO}_2$  (see equation (6) below). Annual precipitation  $P$  ( $\text{m yr}^{-1}$ ) is parameterized via temperature<sup>34</sup>:

$$P = 1.5 \times 2^{(T - T_m)/10} \quad (2)$$

The fraction of precipitation falling as snow, and annual surface melt if any, are computed from  $T$  using a positive-degree-day (PDD)<sup>35</sup> method with coefficient  $0.005 \text{ m per degree-day}$ . A sinusoidal seasonal temperature cycle of amplitude  $0.1 \Delta q_s$  (°C) is assumed, where  $\Delta q_s$  is January-minus-July  $80^\circ\text{S}$  insolation ( $\text{W m}^{-2}$ ). (Very little surface-melt occurs in our simulations, because summer air temperatures remain below freezing everywhere.)

**Modern climatic forcing: sub-ice-shelf oceanic melt.** A new parameterization of oceanic melt rates is used, based on the degree of protection by islands and bays, and distance to ice-shelf edge<sup>17</sup>. Although simple, it captures basic features of other studies, such as rapid melting near edges<sup>14,19,36,37</sup>, and yields reasonable modern shelf distributions. Modern sub-ice melt  $M$  ( $\text{m yr}^{-1}$ ) is

$$M = (1 - z_d) [(1 - z_e) M_p + z_e M_e] + z_d M_d \quad (3)$$

where the ‘deep-ocean’ weighting is

$$z_d = \max[0, \min[1, (h_b - 1400)/200]] \quad (4)$$

and the ‘exposed-shelf’ weighting is

$$z_e = \max[0, \min[1, (A - 80)/30]] e^{-D/100} \quad (5)$$

Here  $\max[x, y]$  indicates the greater of  $x$  and  $y$ , and  $\min[x, y]$  indicates the lesser. The 3 modern oceanic melt rates  $M_p$ ,  $M_e$  and  $M_d$  in equation (3) are for protected, exposed-shelf and deep-ocean areas, respectively, given by  $M_p = 0.1 \text{ m yr}^{-1}$ ,  $M_e = 5 \text{ m yr}^{-1}$ , and  $M_d = 5 \text{ m yr}^{-1}$ . In equations (4) and (5),  $h_b$  is bathymetry (m),  $A$  is the angle (degrees) subtended by the set of all straight lines from the point in question that reach open ocean without encountering land or grounded ice, and  $D$  (km) is the sub-ice distance to the closest open-ocean point. The angle  $A$  is the main way we achieve realistic modern ice-shelf edges. Around most major Antarctic shelf-edges today,  $A$  is  $\sim 90^\circ$  to  $100^\circ$ ; whether this coincidence has a physical basis requires exploration with regional ocean models.

**Past climatic forcing: sea level, temperature and precipitation.** We need to prescribe long-term variations of sub-ice oceanic melt rates, sea level, air temperature and precipitation over the past 5 Myr. On longer timescales, atmospheric  $\text{CO}_2$  levels outside the Plio-Pleistocene range ( $\sim 180$ – $380 \text{ p.p.m.v.}$ ), basal sediment changes, and tectonic uplift or subsidence are probably important, but were probably minor through the Plio-Pleistocene. As mentioned above, our Pleistocene-centric parameterizations may underestimate warmth during the early Pliocene  $\sim 5$ – $3 \text{ Myr}$  ago when  $\text{CO}_2$  levels rose to  $\sim 380 \text{ p.p.m.v.}$  (ref. 9).

Sea-level variations have been dominated by Northern Hemispheric ice volume, and are assumed proportional to deep-sea core  $\delta^{18}\text{O}$  and calibrated as in equation (6). Past variations of Antarctic annual surface temperatures are included in equation (1), proportional to a combination of atmospheric  $\text{CO}_2$  (which is represented by  $\delta^{18}\text{O}$  via sea level in equation (1), since all three are highly correlated in the Pleistocene at least) and the annual  $80^\circ\text{S}$  insolation anomaly. Past variations in precipitation depend on air temperature, just as for modern spatial variations (equation (2)).

**Past climatic forcing: sub-ice-shelf oceanic melt.** The long-term controls of sub-ice-shelf melting are just beginning to be explored<sup>14,19,37</sup>. Here we propose a parameterization based on simple reasoning and sensitivity tests of WAIS retreat since 15 kyr ago. This last deglacial retreat is the only well-documented WAIS variation on  $10^4$ -year time scales. It cannot have been driven by surface mass balance, because Antarctic precipitation has increased, not decreased, and there has been negligible surface melt during this time. Model sensitivity tests show that sea-level rise alone, and/or the influence of warming temperatures on ice viscosity and basal sliding, account for only a small fraction of the observed retreat. Therefore, increases in sub-ice melting must have been key. They could reasonably have been driven either by regional Southern Hemispheric orbital insolation changes, or by global-scale far-field influences. Southern Hemispheric insolation is unlikely to have been the dominant driver, because (1) the summertime  $80^\circ\text{S}$  anomaly from present was small and negative between 15 and 2 kyr ago, and (2) the annual  $80^\circ\text{S}$  anomaly, with minimum at 28.7 kyr ago and maximum at 9.5 kyr ago (ref. 20), would have caused retreat to commence too early (before  $\sim 19 \text{ kyr}$  ago) judging from Ross Sea grounding-line history ( $\sim 10 \text{ kyr}$  ago)<sup>24,25</sup>. This is borne out by sensitivity tests (Supplementary Fig. 5) in which austral insolation is used as the sole driver of sub-ice melt, and results over the past 15,000 years are unreasonable. Realistic retreat is obtained only if sub-ice melt varies in step with far-field forcing.

This suggests that sub-ice melt has been controlled not by local forcing or austral insolation, but by far-field climatic influences that vary in step with Northern Hemispheric glacial–interglacial cycles at least since  $\sim 2.5 \text{ Myr}$  ago. The latter is represented here by a stacked deep-sea-core  $\delta^{18}\text{O}$  record spanning the past 5 Myr (ref. 18). A small influence of austral summer insolation<sup>20</sup> is added to produce minor observed 20-kyr cyclicity during warm events such as MIS 31<sup>21</sup>. First, a weighting index  $w_g$  is defined by

$$w_g = \max[0, \min[2, 1 + \Delta s/85 + \max(0, \Delta q_i/40)]] \quad (6)$$

where  $\delta^{18}\text{O}$  is represented by  $\Delta s$ , the sea-level departure from present (m, scaled to  $\delta^{18}\text{O}$  with last-glacial-maximum 125 m lower than present), and  $\Delta q_i$  is the January  $80^\circ\text{S}$  insolation anomaly from present ( $\text{W m}^{-2}$ ). Sub-ice-melt rates for protected, exposed-shelf and deep-sea areas ( $[M_p, M_e, M_d]$  respectively, in  $\text{m yr}^{-1}$ ) are specified as  $[0, 0, 2]$  for maximum-glacial conditions,  $[0.1, 5, 5]$  for modern, and  $[2, 10, 10]$  for extreme-interglacial conditions. Then the triplet used in equation (3) to determine  $M$  for any past time is

$$[M_p, M_e, M_d] = (1 - w_g) [0, 0, 2] + w_g [0.1, 5, 5] \quad \text{if } 0 \leq w_g < 1 \quad (7)$$

or

$$[M_p, M_e, M_d] = (2 - w_g) [0.1, 5, 5] + (w_g - 1) [2, 10, 10] \quad \text{if } 1 \leq w_g \leq 2 \quad (8)$$

The modern triplet values are chosen to yield reasonable results for today’s Ross and Filchner-Ronne ice shelves. The glacial and warm triplets and the form of  $w_g$  in equation (6) are chosen so that the model just attains full-glacial WAIS extents and complete interglacial collapses in long-term simulations and in Fig. 1. These values cannot be changed by large amounts without substantial degradation of our results.

33. Huybrechts, P. Glaciological modelling of the late Cenozoic East Antarctic Ice Sheet: Stability or dynamism? *Geogr. Ann.* **75**, 221–238 (1993).
34. Huybrechts, P. *Report of the Third EISMINT Workshop on Model Intercomparison* (European Science Foundation, 1998).
35. Marshall, S. J. & Clarke, G. K. C. Ice sheet inception: Subgrid hypsometric parameterization of mass balance in an ice sheet model. *Clim. Dyn.* **15**, 533–550 (1999).
36. Macayeal, D. R. & Thomas, R. H. The effects of basal melting on the present flow of the Ross Ice Shelf, Antarctica. *J. Glaciol.* **32**, 72–86 (1986).
37. Dinniman, M. S., Klinck, J. M. & Smith, W. O. Influence of sea ice cover and icebergs on circulation and water mass formation in a numerical circulation model of the Ross Sea, Antarctica. *J. Geophys. Res.* **112**, C11013, doi:10.1029/2006JC004036 (2007).



# An Early Cretaceous heterodontosaurid dinosaur with filamentous integumentary structures

Xiao-Ting Zheng<sup>1</sup>, Hai-Lu You<sup>2</sup>, Xing Xu<sup>3</sup> & Zhi-Ming Dong<sup>3</sup>

Ornithischia is one of the two major groups of dinosaurs, with heterodontosauridae as one of its major clades. Heterodontosauridae is characterized by small, gracile bodies and a problematic phylogenetic position<sup>1,2</sup>. Recent phylogenetic work indicates that it represents the most basal group of all well-known ornithischians<sup>3</sup>. Previous heterodontosaurid records are mainly from the Early Jurassic period (205–190 million years ago) of Africa<sup>1,3</sup>. Here we report a new heterodontosaurid, *Tianyulong confuciusi* gen. et sp. nov., from the Early Cretaceous period (144–99 million years ago) of western Liaoning Province, China. *Tianyulong* extends the geographical distribution of heterodontosaurids to Asia and confirms the clade's previously questionable temporal range extension into the Early Cretaceous period. More surprisingly, *Tianyulong* bears long, singular and unbranched filamentous integumentary (outer skin) structures. This represents the first confirmed report, to our knowledge, of filamentous integumentary structures in an ornithischian dinosaur.

Dinosauria Owen, 1842

Ornithischia Seeley, 1887

Heterodontosauridae Kuhn, 1966

*Tianyulong confuciusi* gen. et sp. nov.

**Etymology.** The generic name refers to Shandong Tianyu Museum of Nature (STMN; Pingyi, Shandong Province, China), where the specimen is housed; the specific name is dedicated to Confucius, the founder of Confucianism.

**Holotype.** STMN 26-3 is an incomplete skeleton preserving a partial skull and mandible, partial presacral vertebrae, proximal–middle caudal vertebrae, nearly complete right scapula, both humeri, the proximal end of the left ulna, partial pubes, both ischia, both femora, the right tibia and fibula and pes, and filamentous integumentary structures.

**Locality and horizon.** Jianchang County, Liaoning Province, China. Jehol Group, Early Cretaceous<sup>4,5</sup>.

**Diagnosis.** A heterodontosaurid dinosaur distinguishable from other species in having a semi-circular diastema, a barely developed buccal emargination, a dorsally placed mandibular jaw joint well above the level of the dentary tooth row, only one premaxillary tooth (which is caniniform), a distinct gap between the dentary teeth and coronoid process, caudally increasing tooth sizes, and adjacent maxillary and dentary tooth crowns distinctly separate and lacking mesial and distal marginal denticles and ridges on the buccal surfaces.

**Description.** The *T. confuciusi* holotype is estimated to be 70 cm long, with a cranium that is 6 cm long and tail that is 44 cm long based on the proportions of *Heterodontosaurus tucki*<sup>1,6</sup> (Fig. 1a, b; also see Supplementary Information for measurements). The holotype probably represents a subadult individual because the sutural lines between the neural arches and the centra are still visible, whereas the ends of the preserved long bones are all ossified.

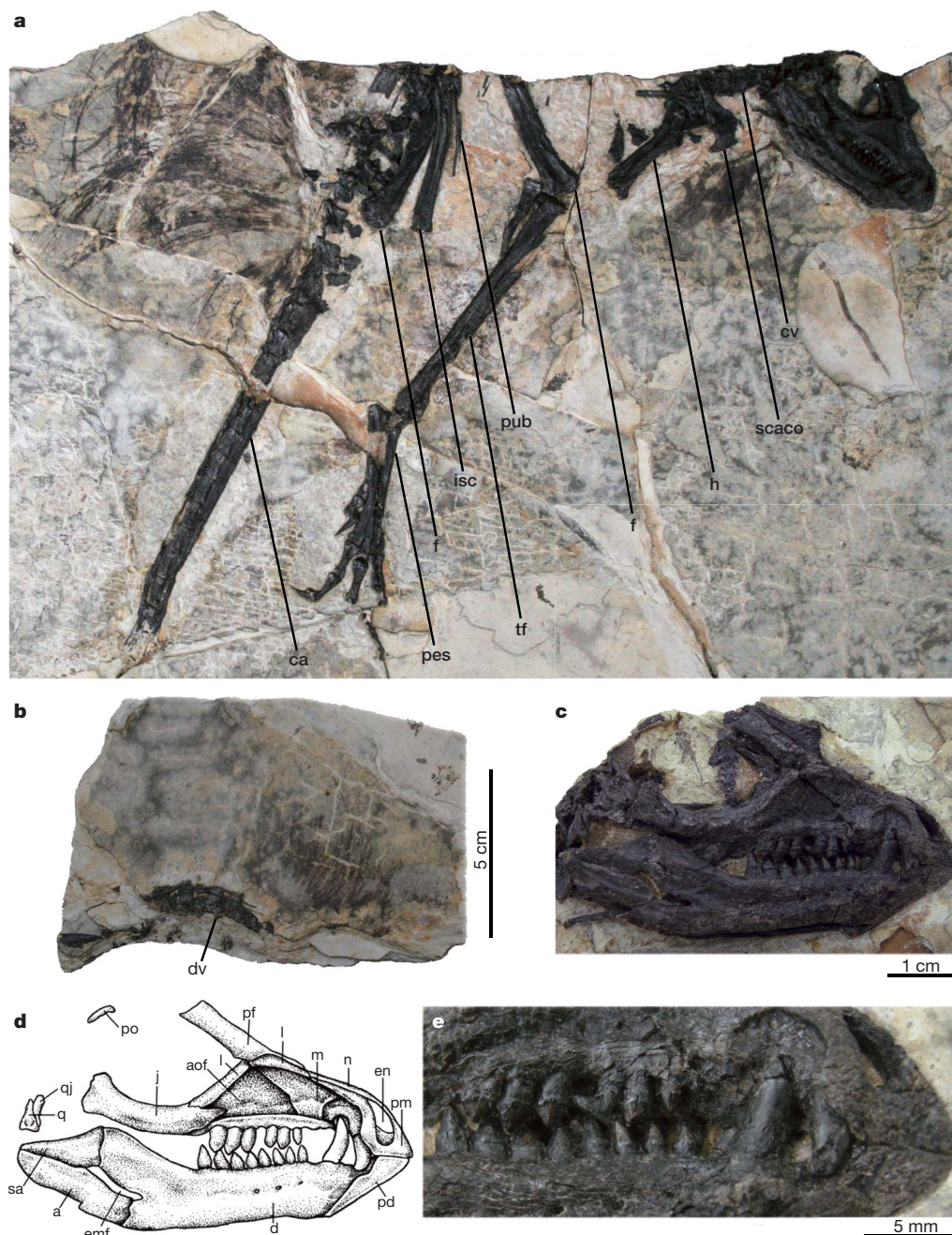
The partial skull, exposed in right lateral view, is missing most of its caudodorsal portion, including the braincase (Fig. 1c, d). A relatively long preorbital portion occupies about 46% of the basal skull length. The ventral edge of the premaxilla lies well below the maxillary tooth row, whereas its dorsal (nasal) process contacts the nasal. A distinct semi-circular diastema is demarcated on the lateral surface across the junction between the premaxilla and the maxilla. Caudal to the diastema, a large, triangular antorbital fossa occupies almost the entire caudal half of the lateral facial area. The long rostradorsal (maxillary) process of the jugal tapers towards its rostral end and inserts between the maxilla and the lacrimal. The lateral mandibular articular condyle of the quadrate is slightly larger than the medial one. The prementary has a smooth, horizontal oral margin of equal length as the oral margin of the premaxilla. The dentary has roughly parallel dorsal and ventral margins, and its horizontal tooth row occupies its rostral two-thirds. A slender external mandibular fenestra is surrounded by the dentary, surangular and angular. The ventral margin of the angular tilts caudodorsally, positioning the jaw joint well above the level of the dentary tooth row. Only one large, caniniform premaxillary tooth exists and protrudes from the caudal end of the oral margin of the premaxilla. The maxillary tooth crowns are triangular in lateral view and lack both denticles along the mesial and distal edges and ridges on the buccal surface. Adjacent crowns are clearly separate from each other. Compared to the maxillary crowns, individual dentary teeth are relatively small and have more pointed apical ends.

The cervical vertebrae are slightly downturned, whereas the dorsals are arched dorsally. Most of the 22 proximal–middle caudals are tightly bound by longitudinally arranged ossified epaxial and hypaxial tendons. The relatively long, straight humerus has a prominent deltopectoral crest. The ischium is straight, compressed mediolaterally, and does not possess tab-shaped obturator process. The postpubis is about half the length of the ischium. The distal end of the femur has an intercondylar extensor groove and an open flexor intercondylar groove. The four preserved metatarsals are closely appressed to each other; metatarsal I is reduced, and its first phalanx does not extend beyond the proximal end of metatarsal II.

Three patches of filamentous integumentary structures are preserved: one ventral to the cervical vertebrae (Fig. 2a), one dorsal to the dorsal vertebrae (Fig. 2b), and the largest dorsal to the proximal caudal vertebrae (Fig. 2c). The filaments in each patch largely parallel each other; those ventral to the body angle about 45° caudally, whereas those dorsal to the body lie almost perpendicular to the long axes of their respective parts of the vertebral column but curve gently. We interpret these as integumentary structures based on their general morphology and distribution along the fossil body. The elongation and hollow nature of these filaments argues against them being subdermal structures, such as collagen fibres<sup>7,8</sup>.

The tail filaments are especially long, with estimated overall lengths of 60 mm (seven times the lengths of the underlying caudal centra) and

<sup>1</sup>Shandong Tianyu Museum of Nature, Lianhuashan Road West, Pingyi, Shandong, 273300, China. <sup>2</sup>Institute of Geology, Chinese Academy of Geological Sciences, 26 Baiwanzhuang Road, Beijing 100037, China. <sup>3</sup>Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences, 142 Xiwai Street, Beijing 100044, China.



**Figure 1 | New heterodontosaurid ornithischian dinosaur *T. confuciusi*.**

**a**, Main slab of the holotype (STMN 26-3). **b**, Broken slab. The scale bar in **b** refers also to **a**. **c**, Close-up of skull and mandible. **d**, Interpretive drawing of skull and mandible. **e**, Close-up of dentition. Abbreviations: a, angular; aof, antorbital fossa; ca, caudal vertebrae; cv, cervical vertebrae; d, dentary;

dv, dorsal vertebrae; emf, external mandibular fenestra; en, external naris; f, femur; h, humerus; isc, ischium; j, jugal; l, lacrimal; m, maxilla; n, nasal; pd, premandible; pf, prefrontal; pm, premaxilla; po, postorbital; pub, pubis; q, quadrate; qj, quadratojugal; scaco, scapulocoracoid; sa, surangular; tf, tibia and fibula.

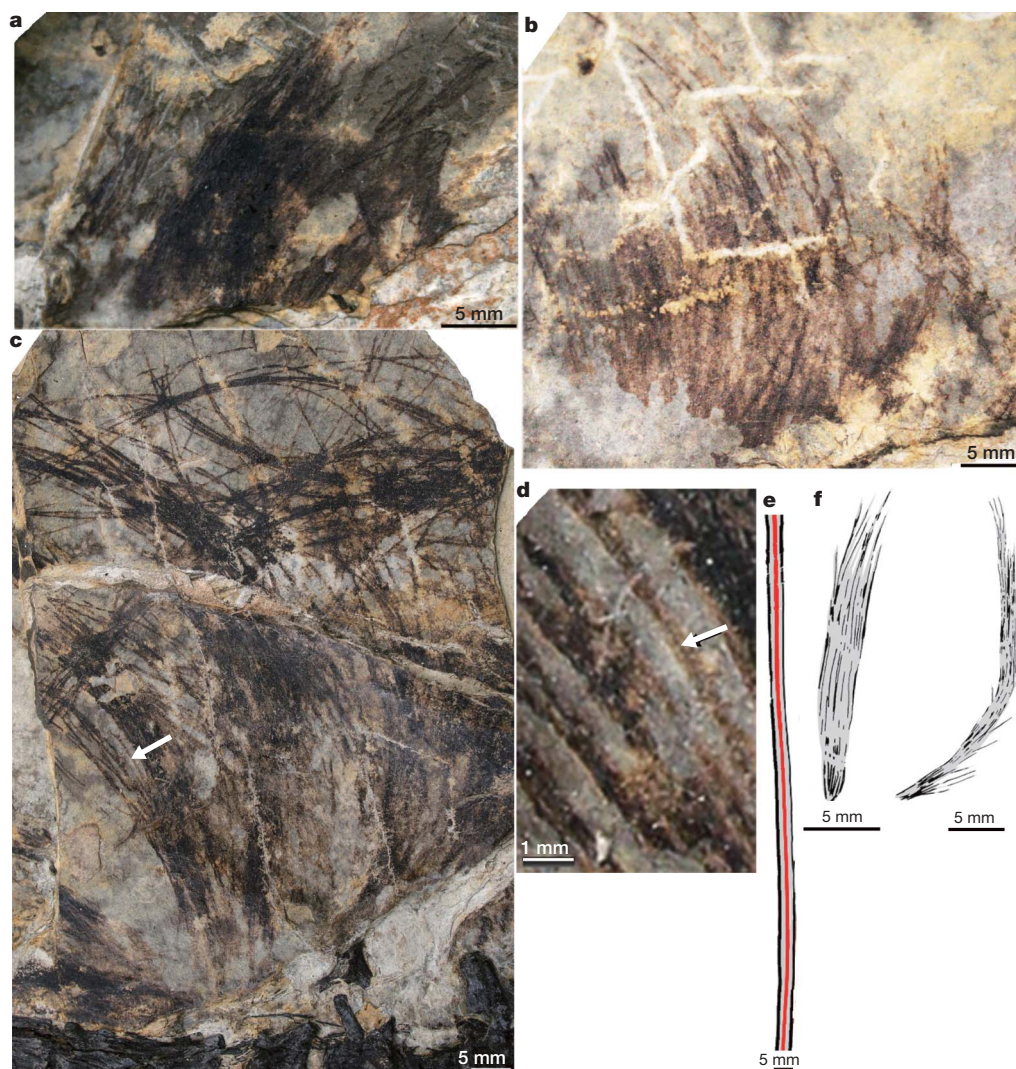
widths of 0.4 mm. Although their bases are not clearly preserved, all the filaments seem to be singular, showing no evidence of branching. The filaments appear to have been relatively rigid because none are wavy or bent and they do not deviate along their lengths from otherwise gentle curvatures. The widths of all filaments appear similar in each patch, which strongly suggests that they have roughly constant radial dimensions and therefore circular cross sections. Although largely overlapping, several isolated, partial filaments on the tail also display dark, longitudinal 'stripes', implying that each had a tubular structure (Fig. 2d).

Given the stronger evidence for a tubular structure and their apparent rigidity, the filamentous integumentary structures of *Tianyulong* are more similar to those preserved on the tail of a specimen of *Psittacosaurus*<sup>9</sup> than to the 'protofeather' structures in both

avian and non-avian theropods<sup>10,11</sup>. However, the integumentary structures in *Psittacosaurus* are preserved only on the tail, and not distributed more widely along the body as in *Tianyulong*. Also, those in *Psittacosaurus* are more rigid and more widely separated from each other than in *Tianyulong*. The central tail feathers of the basal avialan *Epidexipteryx*<sup>12</sup> (Fig. 2e), the basal pygostylian (all birds with shortened tails) *Confuciusornis*<sup>13</sup> and some enantiornithean (one of the two major clades of early birds, which became extinct by the end of Cretaceous period) birds<sup>14,15</sup> are also elongate, and represent a feather type unknown in extant birds. Unlike the structures in *Tianyulong*, however, they are ribbon-like and branch at their distal ends<sup>10</sup>.

Among the 'protofeathers' of theropods, the integumentary growths of *Tianyulong* are more similar to those of *Sinosauropteryx*<sup>16</sup> than to





**Figure 2 | Integumentary structures of *T. confuciusi* holotype (STMN 26-3) and others for comparison. a, Ventral to the cervical vertebral series. b, Dorsal to the dorsal vertebral series. c, Dorsal to the proximal–middle caudal vertebral series. d, Close-up of c. Arrows in c and d point to single filament exhibiting a clear, dark, midline ‘stripe’. e, Schematic of long, central tail feather of *Epidexipteryx* (after ref. 12). f, Two types of integumentary filaments of *Sinornithosaurus* (after ref. 17).**

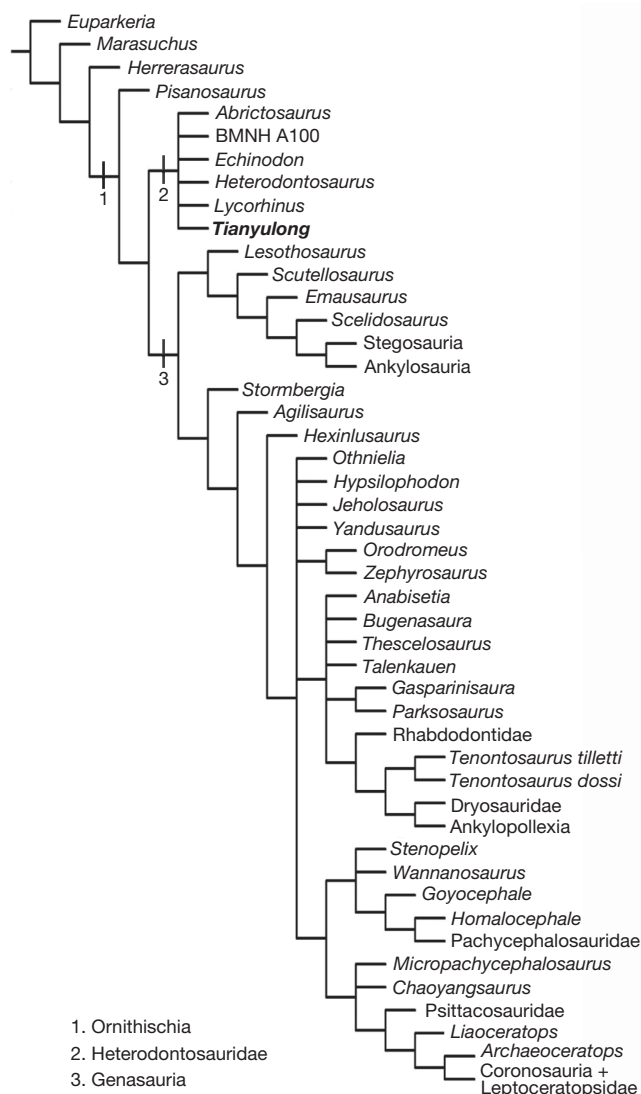
those of *Sinornithosaurus*<sup>17</sup> (Fig. 2f) and others. In both *Tianyulong* and *Sinosauropteryx*, the filamentous structures are singular and unbranched, although the ‘protofeathers’ in the similar-sized *Sinosauropteryx* are much shorter and more slender than the caudal structures in *Tianyulong*, with lengths of less than 30 mm and widths of less than 0.2 mm<sup>16</sup>. In many of the theropods, the ‘protofeathers’ are more or less bundled together (Fig. 2f, left); in others, they exhibit a clear branching pattern (Fig. 2f, right). Overall, they appear to have been much more pliable<sup>10,11,16</sup>.

Phylogenetic analysis recovers *Tianyulong* as a member of Heterodontosauridae, which, in turn, is the basalmost coherent group among all well-known ornithischians<sup>3</sup> (Fig. 3; also see Supplementary Information). Shared derived features of heterodontosaurids include: ventral deflection of the oral margin of the premaxilla ventral to the maxillary tooth row, a large premaxilla–maxilla diastema at least one crown in length, quadrate with larger lateral than medial mandibular condyle, and a large, caudally placed caniniform premaxillary tooth; all these features are present in *Tianyulong*. Besides three heterodontosaurid genera from the Early Jurassic of southern Africa<sup>1,18</sup>, other reports include one genus from the Early Cretaceous of England<sup>19</sup>, as well as fragmentary Middle and Late Jurassic–Early Cretaceous material from Europe<sup>19,20</sup> and the Late Jurassic of North America<sup>21</sup>. Thus, *Tianyulong* expands the distribution of heterodontosaurids into Asia and confirms a long temporal extension for the clade into the Early Cretaceous, bolstering a heterodontosaurid interpretation for the European and North American reports.

The discovery of filamentous integumentary structures in *Tianyulong* provides an unprecedented phylogenetic extension of archosaurian dermal structures, previously reported only in derived theropodan saurischian dinosaurs, to Ornithischia. It also expands the known diversity of elaborate integumentary structures in Mesozoic reptiles, such as those in the diapsid *Longisquama*<sup>22,23</sup>, pterosaurs<sup>24,25</sup>, non-avian coelurosaurian theropods<sup>10,11</sup> and birds<sup>10,11</sup>, to ornithischians. However, the homology of these structures across taxa remains poorly understood. In terms of their relationships to feathers, both convergence<sup>23,24,26</sup> and homology<sup>4,11,25,27</sup> have been proposed to explain the elongate dorsal projections of *Longisquama* and the short, hair-like structures of pterosaurs; homology is favoured for the structures of theropods<sup>4,11,28</sup> because they share features (such as branching) known elsewhere only in feathers. The unique filaments of *Tianyulong* add more complexity to the issue of feather origins. Homology of the structures in *Tianyulong* and theropods is far from obvious with present data, but cannot be precluded. Although based on negative evidence, the derived position within the Theropoda for the known appearance of ‘protofeathers’ indicates that earlier theropods lacked integumentary structures, implying in turn that the common ancestor of theropods and ornithischians also lacked such structures and that their appearances in each clade were convergent.

When homology with ‘protofeathers’ is hypothesized, a possible scenario is that the elongate, singular, cylindrical filaments of *Tianyulong* represent elaborations of the initial stage in development models of the origin and evolution of feathers<sup>28,29</sup>. The basal position





**Figure 3 | Phylogenetic relationships of *T. confuciusi* and Heterodontosauridae.** Cladistic analysis recovers *Tianyulong* as a member of Heterodontosauridae, which is itself a basal clade of ornithischian dinosaurs. This is based on a strict consensus of 8 most parsimonious trees with a tree length of 485, consistency index of 0.501 and retention index of 0.730. This analysis was performed by a sectorial search of TNT (tree analysis using new technology)<sup>30</sup> (see Supplementary Information).

of Heterodontosauridae within Ornithischia then suggests that such early-stage structures were present in the earliest dinosaurs (before or at the Saurischia–Ornithischia split) and inherited by basal members of each group. Later ornithischians and non-theropod saurischians for which skin impressions are known would therefore represent secondary losses of these structures at as-yet undetermined systematic positions. The structures in *Tianyulong*, as well as the bristle-like structures of the basal ceratopsian *Psittacosaurus*<sup>9</sup>, which have frequently been perceived as nonhomologous with the filamentous structures of theropods, may truly be homologous, albeit derived in structure compared to those of theropods.

Received 7 December 2008; accepted 28 January 2009.

1. Norman, D. B., Sues, H.-D., Witmer, L. M. & Coria, R. A. in *The Dinosauria* 2nd edn (eds Weishampel, D. B., Dodson, P. & Osmólska, H.) 393–412 (Univ. California Press, 2004).
2. Xu, X., Forster, C. A., Clark, J. M. & Mo, J. A basal ceratopsian with transitional features from the Late Jurassic of northwestern China. *Proc. R. Soc. Lond. B* **273**, 2135–2140 (2006).

3. Butler, R. J., Upchurch, P. & Norman, D. B. The phylogeny of the ornithischian dinosaurs. *J. Syst. Palaeontol.* **6**, 1–40 (2008).
4. Zhou, Z., Barrett, P. M. & Hilton, J. An exceptionally preserved Lower Cretaceous ecosystem. *Nature* **421**, 807–814 (2003).
5. Ji, Q. et al. *Mesozoic Jehol Biota of Western Liaoning, China* (Geological Publishing House, 2004).
6. Weishampel, D. B. & Witmer, L. M. in *The Dinosauria* 1st edn (eds Weishampel, D. B., Dodson, P. & Osmólska, H.) 486–497 (Univ. California Press, 1990).
7. Lingham-Soliar, T., Feduccia, A. & Wang, X. A new Chinese specimen indicates that 'protofeathers' in the Early Cretaceous theropod dinosaur *Sinosauropteryx* are degraded collagen fibres. *Proc. R. Soc. Lond. B* **274**, 1823–1829 (2007).
8. Feduccia, A., Lingham-Soliar, T. & Hinchliffe, J. R. Do feathered dinosaurs exist? Testing the hypothesis on neontological and paleontological evidence. *J. Morphol.* **266**, 125–166 (2005).
9. Mayr, G., Peters, D. S., Plodowski, G. & Vogel, O. Bristle-like integumentary structures at the tail of the horned dinosaur *Psittacosaurus*. *Naturwissenschaften* **89**, 361–365 (2002).
10. Zhang, F., Zhou, Z. & Dyke, G. J. Feathers and 'feather-like' integumentary structures in Liaoning birds and dinosaurs. *Geol. J.* **41**, 395–404 (2006).
11. Xu, X. Feathered dinosaurs from China and the evolution of major avian characters. *Integ. Zool.* **1**, 4–11 (2006).
12. Zhang, F. et al. A bizarre Jurassic maniraptoran from China with elongate ribbon-like feathers. *Nature* **455**, 1105–1108 (2008).
13. Zhang, F., Zhou, Z. & Benton, M. J. A primitive confuciusornithid bird from China and its implications for early avian flight. *Sci. China Ser. D* **51**, 625–639 (2008).
14. Zhang, F. & Zhou, Z. A primitive enantiornithine bird and the origin of feathers. *Science* **290**, 1955–1960 (2000).
15. Zheng, X., Zhang, Z. & Hou, L. A new enantiornithine bird with four long rectrices from the Early Cretaceous of northern Hebei, China. *Acta Geol. Sin.-Engl.* **81**, 703–708 (2007).
16. Currie, P. J. & Chen, P. J. Anatomy of *Sinosauropteryx prima* from Liaoning, northeastern China. *Can. J. Earth Sci.* **38**, 1705–1727 (2001).
17. Xu, X., Zhou, Z.-H. & Prum, R. O. Branched integumental structures in *Sinornithosaurus* and the origin of feathers. *Nature* **410**, 200–204 (2001).
18. Butler, R. J., Porro, L. B. & Norman, D. B. A juvenile skull of the primitive ornithischian dinosaur *Heterodontosaurus tucki* from the 'Stormberg' of southern Africa. *J. Vertebr. Paleontol.* **28**, 702–711 (2008).
19. Naish, D. & Martill, D. M. Dinosaurs of Great Britain and the role of the Geological Society of London in their discovery: Ornithischia. *J. Geol. Soc. Lond.* **165**, 613–623 (2008).
20. Sánchez-Hernández, B., Benton, M. J. & Naish, D. Dinosaurs and other fossil vertebrates from the Late Jurassic and Early Cretaceous of the Galve area, NE Spain. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **249**, 180–215 (2007).
21. Norman, D. B. & Barrett, P. M. Ornithischian dinosaurs from the Lower Cretaceous (Berriasian) of England. *Spec. Pap. Palaeontol.* **68**, 161–189 (2002).
22. Sharov, A. G. An unusual reptile from the Lower Triassic of Fergana. *Paleontol. J.* **1970**, 112–116 (1970).
23. Voigt, S. et al. Feather-like development of Triassic diapsid skin appendages. *Naturwissenschaften* **96**, 81–86 (2009).
24. Wang, X., Zhou, Z., Zhang, F. & Xu, X. A nearly completely articulated rhamphorhynchoid pterosaur with exceptionally well-preserved wing membranes and 'hairs' from Inner Mongolia, northeast China. *Chin. Sci. Bull.* **47**, 226–230 (2002).
25. Ji, Q. & Yuan, C.-X. Discovery of two kinds of protofeathered pterosaurs in the Mesozoic Daohugou Biota in the Ningcheng region and its stratigraphic and biologic significances. *Geol. Rev.* **48**, 221–224 (2002).
26. Reisz, R. R. & Sues, H. D. The 'feathers' of Longisquama. *Nature* **408**, 428 (2000).
27. Jones, T. D. et al. Nonavian feathers in a late Triassic archosaur. *Science* **288**, 2202–2205 (2000).
28. Prum, R. O. & Brush, A. H. The evolutionary origin and diversification of feathers. *Q. Rev. Biol.* **77**, 261–295 (2002).
29. Yu, M., Wu, P., Wideltz, R. B. & Chuong, C.-M. The morphogenesis of feathers. *Nature* **420**, 308–312 (2002).
30. Goloboff, P. A., Farris, J. S. & Nixon, K. C. TNT, a free program for phylogenetic analysis. *Cladistics* **24**, 774–786 (2008).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank P.-J. Chen for arranging the study, Y.-Q. Zhang for preparing the specimen, J. D. Harris for improving the manuscript, and L.-D. Xing for drawing Fig. 1d. H.-L.Y. was supported by grants from the National Natural Science Foundation of China (40672007), the 973 Project and the Basic Outlay of Scientific Research Work from the Ministry of Science and Technology, and the Hundred Talents Project of the Ministry of Land and Resources of China.

**Author Contributions** X.-T.Z. designed the project. H.-L.Y., X.X. and Z.-M.D. performed the research. H.-L.Y. wrote the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to H.-L.Y. ([youhailu@gmail.com](mailto:youhailu@gmail.com)).

# Population genomics of domestic and wild yeasts

Gianni Liti<sup>1\*</sup>, David M. Carter<sup>2\*</sup>, Alan M. Moses<sup>2,3</sup>, Jonas Warringer<sup>4</sup>, Leopold Parts<sup>2</sup>, Stephen A. James<sup>5</sup>, Robert P. Davey<sup>5</sup>, Ian N. Roberts<sup>5</sup>, Austin Burt<sup>6</sup>, Vassiliki Koufopanou<sup>6</sup>, Isheng J. Tsai<sup>6</sup>, Casey M. Bergman<sup>7</sup>, Douda Bensasson<sup>7</sup>, Michael J. T. O'Kelly<sup>8</sup>, Alexander van Oudenaarden<sup>8</sup>, David B. H. Barton<sup>1</sup>, Elizabeth Bailes<sup>1</sup>, Alex N. Nguyen Ba<sup>3</sup>, Matthew Jones<sup>2</sup>, Michael A. Quail<sup>2</sup>, Ian Goodhead<sup>2†</sup>, Sarah Sims<sup>2</sup>, Frances Smith<sup>2</sup>, Anders Blomberg<sup>4</sup>, Richard Durbin<sup>2\*</sup> & Edward J. Louis<sup>1\*</sup>

Since the completion of the genome sequence of *Saccharomyces cerevisiae* in 1996 (refs 1, 2), there has been a large increase in complete genome sequences, accompanied by great advances in our understanding of genome evolution. Although little is known about the natural and life histories of yeasts in the wild, there are an increasing number of studies looking at ecological and geographic distributions<sup>3,4</sup>, population structure<sup>5–8</sup> and sexual versus asexual reproduction<sup>9,10</sup>. Less well understood at the whole genome level are the evolutionary processes acting within populations and species that lead to adaptation to different environments, phenotypic differences and reproductive isolation. Here we present one- to fourfold or more coverage of the genome sequences of over seventy isolates of the baker's yeast *S. cerevisiae* and its closest relative, *Saccharomyces paradoxus*. We examine variation in gene content, single nucleotide polymorphisms, nucleotide insertions and deletions, copy numbers and transposable elements. We find that phenotypic variation broadly correlates with global genome-wide phylogenetic relationships. *S. paradoxus* populations are well delineated along geographic boundaries, whereas the variation among worldwide *S. cerevisiae* isolates shows less differentiation and is comparable to a single *S. paradoxus* population. Rather than one or two domestication events leading to the extant baker's yeasts, the population structure of *S. cerevisiae* consists of a few well-defined, geographically isolated lineages and many different mosaics of these lineages, supporting the idea that human influence provided the opportunity for cross-breeding and production of new combinations of pre-existing variations.

The baker's yeast *S. cerevisiae* has had a long association with human activity<sup>11</sup>, leading to the idea that its use in fermentation led to its domestication. Two domestication events have been suggested, one for sake strains and one for wine<sup>12</sup>. In contrast, its closest relative, *S. paradoxus*, has never been associated with human activity and is found globally, sometimes in the same locations as *S. cerevisiae*<sup>3,4</sup>. A preliminary comparison within the *Saccharomyces sensu stricto* group exhibited extensive variation between *S. paradoxus* populations on different continents but limited variation among *S. cerevisiae* isolates and no correlation with geographic location<sup>8</sup>.

Here we report nearly complete genome sequences of *S. cerevisiae* and *S. paradoxus* from a large variety of sources and locations (Supplementary Tables 1 and 2). The *S. cerevisiae* strains included the reference strain S288c plus other lab, pathogenic, baking, wine, food spoilage, natural fermentation, sake, probiotic and plant isolates.

The *S. paradoxus* isolates were mostly from oak tree bark from the three recognized populations<sup>6,8,13</sup> as well as Siberia, Hawaii and the previously designated *Saccharomyces cariocanus*<sup>14</sup>. There is overlap among the general geographic sources of isolates from both species. The majority of strains were sequenced using Sanger sequencing on ABI 3730 DNA sequencers (Applied Biosystems). For some strains, sequence was obtained using an Illumina Genome Analyzer. Most strains were covered to one- to fourfold depth with a few covered more extensively (Supplementary Table 3). The sequence reads, assemblies, alignments, a BLAST tool and a genome browser are all publicly available<sup>15</sup>.

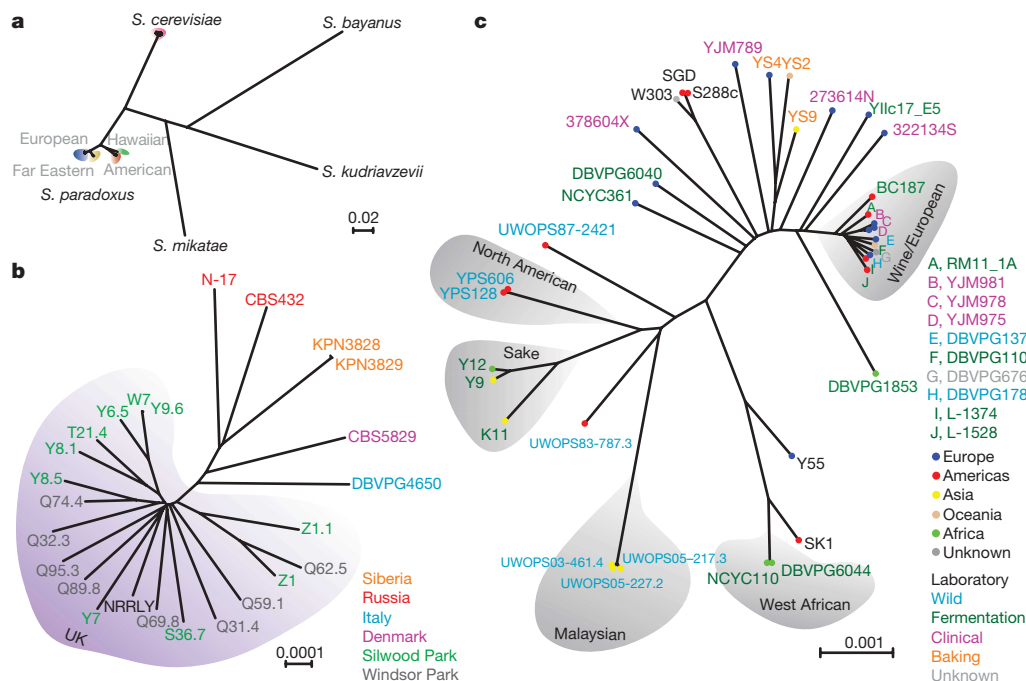
We identified 235,127 high-quality single nucleotide polymorphisms (SNPs) and 14,051 nucleotide insertions or deletions (indels) in the *S. cerevisiae* nuclear genome, and 623,287 SNPs and 25,267 indels in *S. paradoxus*. Our S288c sequence differs from the reference genome by 498 high-quality, unambiguous SNPs (Supplementary Fig. 1). For 480 SNPs our S288c sequence is supported by other strains whereas the reference has no support, and for 18 SNPs the reference sequence is supported by other strains whereas ours is not. Many of the 480 SNPs are likely to represent errors in the reference sequence (Supplementary Table 4). The reference sequence for the type strain of *S. paradoxus*<sup>16</sup> was not complete, so we sequenced the type strain CBS432 to 4.3-fold coverage with an ABI 3730 sequencer and to 80-fold coverage with the Illumina Genome Analyzer.

Sequence surveys allow novel sequences not found in the reference genome to be identified. The proportions of unplaced reads for each strain are shown in Supplementary Table 2. We found 38 new hypothetical open reading frames (ORFs) in these sequences that are likely to be real. These ORFs are present in more than one strain (Supplementary Fig. 2), with some specific to a single lineage, such as the hypothetical protein 5 (Supplementary Information) in the West African lineage, which contains a conserved methyltransferase domain. Much of the unplaced material is subtelomeric. This is in contrast to a genome-wide analysis of copy number based on the numbers of reads of each strain aligning to each gene in the reference sequence, which showed very little significant copy number variation outside the ribosomal DNA (rDNA) region (Supplementary Information).

Neighbour-joining phylogenetic trees based on pairwise SNP differences in the alignments were generated (Fig. 1 and Supplementary Fig. 3). The *S. paradoxus* strains fall into the three previously described populations, plus one isolate from Hawaii. Most of the SNPs in *S. paradoxus* are private polymorphisms within each population,

<sup>1</sup>Institute of Genetics, Queen's Medical Centre, University of Nottingham, Nottingham NG7 2UH, UK. <sup>2</sup>Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1HH, UK. <sup>3</sup>Department of Cell & Systems Biology, University of Toronto, Ontario M5S 2J4, Canada. <sup>4</sup>Department of Cell and Molecular Biology, Lundberg Laboratory, University of Gothenburg, Medicinaregatan 9c, 41390 Gothenburg, Sweden. <sup>5</sup>National Collection of Yeast Cultures, Institute of Food Research, Norwich Research Park, Colney, Norwich NR4 7UA, UK. <sup>6</sup>Division of Biology, Imperial College London, Silwood Park, Ascot SL5 7PY, UK. <sup>7</sup>Faculty of Life Sciences, University of Manchester, Manchester M13 9PT, UK. <sup>8</sup>Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. <sup>†</sup>Present address: School of Biological Sciences, University of Liverpool, Liverpool L69 3BX, UK.

\*These authors contributed equally to this work.



**Figure 1 | *Saccharomyces* phylogenomics.** **a**, Neighbour-joining trees based on SNP differences of *S. cerevisiae* and *S. paradoxus* strains sequenced in this project, using *Saccharomyces mikatae*, *Saccharomyces kudriavzevii* and *Saccharomyces bayanus* as out-groups. **b**, Close-up of the European

*S. cerevisiae* strains with clean lineages highlighted in grey; colour indicates source (name) and geographic origin (dots). Scale bars indicate frequencies of base-pair differences.

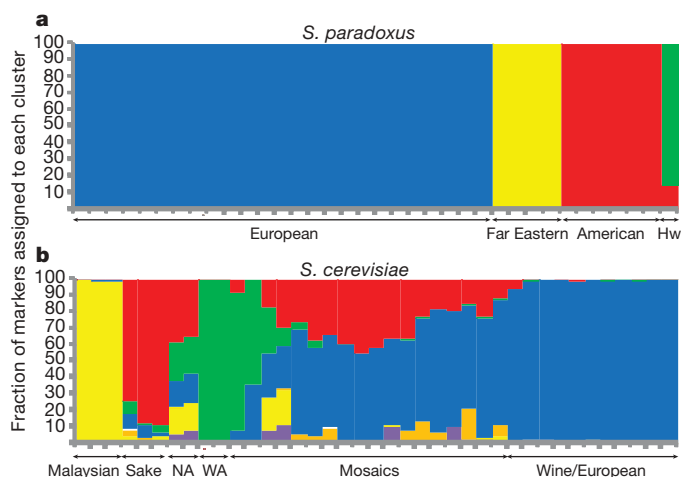
resulting in a clear separation of the three populations<sup>17</sup> (Fig. 2a). The European population was sampled extensively, which provided a picture of within-population structure (Fig. 1b).

The *S. cerevisiae* population structure is more complex. There are five lineages that exhibit the same phylogenetic relationship across their entire genomes, which we consider to be 'clean' non-mosaic lineages (Fig. 1c). These are strains from Malaysia, West Africa, sake and related fermentations (labelled 'Sake' in Fig. 1c), North America, and a large cluster of mixed sources containing many European and wine strains ('Wine/European'). The remaining strains are on long branches between the Wine/European cluster and the other four clean lineages. Although some lineages correspond to geographic origin, such as those from North America and Malaysia, many closely related strains are from widely separated locations. This mixed architecture could be due to human traffic in yeast strains and subsequent recombination between them. Analysis with Structure is consistent with separate populations for the West African, Malaysian, Sake and Wine/European lineages (Fig. 2b). The North American isolates share some polymorphisms with all four separate populations, whereas the rest of the strains share polymorphisms with the European lineage and at least one other population. Analysis of SNP distributions (Supplementary Table 5) is consistent with the neighbour-joining tree phylogeny (Fig. 1c) and the Structure analysis (Fig. 2b). Each clean lineage is monomorphic for the majority of segregating sites, whereas the mosaics are polymorphic for the majority of sites.

Phylogenetic trees constructed for individual chromosomes or smaller segments (Supplementary Fig. 4) demonstrate the mosaic nature of these genomes, as do segmental comparisons (Supplementary Fig. 5). For example, the laboratory strains SK1 and Y55 appear to be the result of recent crosses between the West African lineage and the European lineage (Supplementary Fig. 5b). Similarly, W303 is a recent cross between the reference S288c lineage and one or more other lineages. Different segments of the mosaics fall into different locations in the neighbour-joining tree (Supplementary Fig. 4). The recently sequenced clinical derivative YJM789 (ref. 18) is another example. This complex population structure of *S. cerevisiae* is reported in a similar study<sup>19</sup> and is consistent with five well-delineated lineages,

two of which contain isolates used in fermentation industries<sup>12</sup>, plus a number of recombinant strains, many of which are also used for fermentation. Phenotypic profiling (see below), and analyses of rDNA repeat unit variation (Supplementary Fig. 6) and Ty element abundance (Supplementary Fig. 7 and Supplementary Table 6), produce results consistent with this overall picture of the *S. cerevisiae* population structure.

It is unlikely that the entire sequence space of *S. cerevisiae* has been sampled. It is clear that segments from many of the mosaic strains are



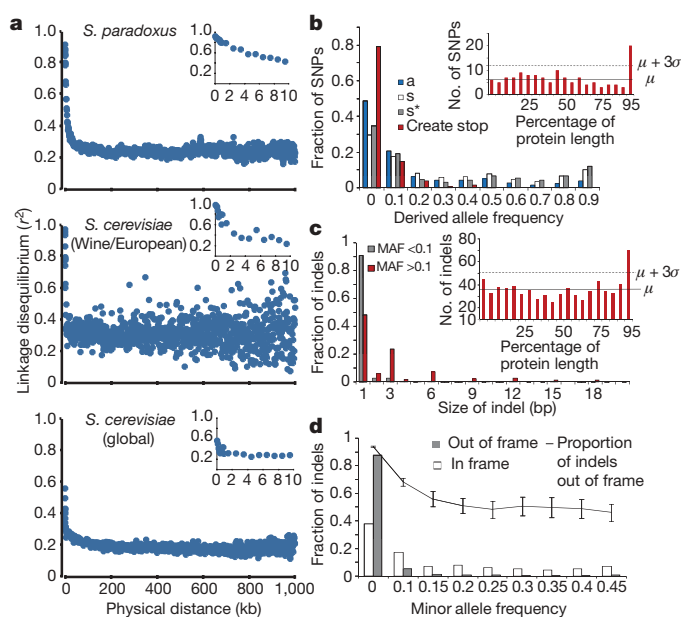
**Figure 2 | *Saccharomyces* population structure.** **a**, Inference of population structure using the program Structure (version 2.1) on *S. paradoxus* (markers: 7,544 SNPs with >30 strains passing neighbourhood quality standard), assuming  $K = 6$  subpopulations and correlated allele frequencies, linkage model based on marker distances in base pairs, 15,000-iteration burn in, and 5,000 iterations of sampling. Each mark on the x axis represents one strain, and the blocks of colour represent the fraction of the genetic material in each strain assigned to each cluster. Hw, Hawaiian isolate. **b**, As in **a**, but for *S. cerevisiae* (markers: 3,413 SNPs with >30 strains passing neighbourhood quality standard). NA, North America; WA, West Africa.



not related to any of the five clean lineages and are probably derived from lineages that are yet to be determined or no longer exist. One-quarter (24%) of SNPs are found only in the mosaics (Supplementary Table 5), which provides a measure of the unsampled *S. cerevisiae* species space.

Sequence variability was quantified using the average pairwise divergence within a population ( $\theta_\pi$ ) and the proportion of polymorphic sites ( $\theta_s$ )<sup>10</sup>. We estimated these parameters for various populations (Supplementary Table 7). Both  $\theta_\pi$  and  $\theta_s$  are about 0.001 in the UK population of *S. paradoxus*. The Wine/European cluster of *S. cerevisiae* has approximately the same level of diversity. In both the global and Wine/European samples of *S. cerevisiae*, Tajima's *D* (ref. 20) is significantly negative, indicating an excess of singleton polymorphisms, which may be a consequence of our sampling strategy. By contrast, the UK sample of *S. paradoxus* from a single population has a positive Tajima's *D*, although not significantly, indicating a relative abundance of mid-frequency polymorphisms. Linkage disequilibrium differs between samples (Fig. 3a). For *S. paradoxus*, linkage disequilibrium declines smoothly with distance, decaying to half its maximum value at about 9 kb, as previously reported<sup>10</sup>. For both *S. cerevisiae* samples, the linkage disequilibrium decays much faster, with a half maximum at 3 kb or less. This implies more recombination in *S. cerevisiae*, perhaps due to more opportunities for strains to mate and recombine.

Patterns of variation can reveal evidence of natural selection. As expected for weakly deleterious mutations, the derived allele frequencies



**Figure 3 | Population genomics: variation and selection.** **a**, Linkage disequilibrium as a function of distance averaged over one kilobase (kb), expressed in terms of correlation coefficient,  $r^2$ . Insets show the decline in linkage disequilibrium over the first 10 kb. Details are shown in Supplementary Table 7. **b**, Derived allele frequencies of SNPs in coding regions. Amino-acid-changing SNPs (labelled 'a') show an excess of low frequencies in comparison with synonymous SNPs ('s'). Synonymous SNPs in genes with strong codon bias ('s\*') are in excess at low and high frequencies. SNPs that create stop codons ('create stop') show skew to low frequencies. Inset is the number of mutations occurring over the length of the protein, exceeding three standard deviations ( $\sigma$ ) from the mean ( $\mu$ ) in the C terminus. **c**, Distribution of sizes of indel polymorphisms in coding regions. High-frequency indels (>10%, red) more often occur in multiples of three than do low-frequency indels (grey). Inset is as for **b**. **d**, Frequency distribution of indels in coding regions. Out-of-frame indels (grey) show excess at low frequencies relative to in-frame indels (open). The proportion of out-of-frame indels decreases as frequency increases. Error bars represent the standard error of the proportion. Numbers of observations for each bin: 0,  $n = 2,910$ ; 0.1,  $n = 184$ ; 0.15,  $n = 68$ ; 0.2,  $n = 52$ ; 0.25,  $n = 29$ ; 0.3,  $n = 29$ ; 0.35,  $n = 36$ ; 0.4,  $n = 29$ ; 0.45,  $n = 40$ .

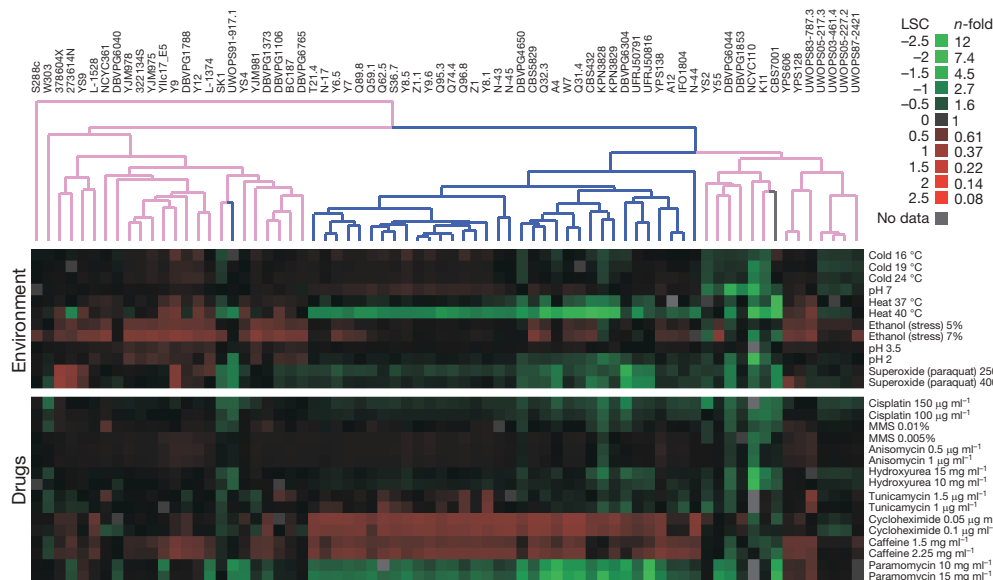
(DAF; see Supplementary Information) for non-synonymous polymorphism are lower than synonymous polymorphism (Fig. 3b). For polymorphisms with DAF <20%, there were 0.86 amino-acid-changing polymorphisms for each silent one. In contrast, for those with DAF >20% this ratio was 0.34, indicating that at least 61% ( $1 - 0.34/0.86$ ) of the 24,418 amino-acid-changing polymorphisms with DAF <20% are deleterious. Similar calculations (Supplementary Information) indicate that 27% of non-coding polymorphisms with DAF <20% are deleterious (Supplementary Fig. 8a). We also performed McDonald–Kreitman tests<sup>21</sup> on 1,105 genes for which we had enough statistical power. No evidence for positive selection after a multiple testing correction was found (Supplementary Fig. 8b). These analyses assumed that synonymous polymorphisms are neutral. However, we found an excess of polymorphism at both low and high frequency (Fig. 3b) in genes with high codon bias (codon adaptation index >0.6). Further analysis (Supplementary Information) indicates that codon bias in *S. cerevisiae* is maintained by both purifying and positive selection, as suggested by the mutation-selection-drift model<sup>22</sup>.

A previous genome-wide study in *Arabidopsis*<sup>23</sup> reported a large number of seemingly highly deleterious alleles. We found 134 mutations that were predicted to introduce stop codons (Fig. 3b), including five in genes previously reported to be essential in S288c (Supplementary Information). These mutations showed a skewed frequency distribution and were enriched in the C termini (the final 5% of proteins; Fig. 3b inset).

This data set allowed the consideration of insertions and deletions (Fig. 3c). We identified 3,870 indels in the coding regions of the *S. cerevisiae* population. Of these, 731 had minor allele frequency (MAF) greater than 10%. We also found 657 indels (72 with MAF >10%) in genes identified as essential (Supplementary Information). Indels with MAF >10% predicted to cause frame shifts were enriched in the C-terminal 5% of the protein (Fig. 3c, inset). The proportion of frame-shift to in-frame indels decreases strongly as a function of MAF (Fig. 3d). For example, at MAF >15% there are 1.0 out-of-frame indels for every in-frame indel, in comparison with 15.5 at MAF <10%. We estimate that 93% ( $1 - 1.0/15.5$ ) of the 2,949 out-of-frame indels with MAF <10% are deleterious.

All strains were subjected to high-throughput phenotypic analysis under multiple conditions (Fig. 4 and Supplementary Fig. 9). Growth curves were sampled (>250 time points) over three days and the relevant growth variables—lag (adaptation), rate (slope) and efficiency (maximum density)—were extracted<sup>24</sup>, providing roughly 200 phenotypic traits. The phenotypic variation allowed clustering of strains. There is a high qualitative overlap between the phenotypic clustering and the phylogenies based on SNPs (Figs 1 and 4). Also, the correlation between genotypic and phenotypic similarity within *S. cerevisiae* is surprisingly good (Spearman's rank test: correlation coefficient, 0.30;  $P = 10^{-26}$ ) given that conventional phenotypic taxonomy generally fails even to resolve the *Saccharomyces sensu stricto* species. No individual environment determined the overall correlation between genotype and phenotype.

The *S. paradoxus* strains were well separated from the *S. cerevisiae* strains (Fig. 4), except for the Hawaiian isolate. The phenotypes separating the two species most clearly ( $P < 10^{-9}$ ) were strong *S. paradoxus* resistance to cycloheximide and sensitivity to paramomycin, heat and copper (Supplementary Fig. 10a). The *S. cerevisiae* isolates fell into two groups (Fig. 4). One contains most of the Wine/European and Sake lineages and most of the long-branch recombinants, whereas the other mainly consists of the North American, Malaysian and African lineages. The main phenotypic characteristic separating these groups is rapid growth (short lag and steep slope in rate,  $P < 10^{-4}$ ) for the Wine/European lineages and the mosaics, which could be advantageous for the fermentation processes in which many of these strains are used (Supplementary Fig. 10b). Despite genomic variation, *S. paradoxus* strains (excluding the Hawaiian isolate) show 38% lower phenotypic variation than *S. cerevisiae* strains ( $P = 0.002$ ). In *S. cerevisiae*, the phenotypic variance is as high among the clean lineages as among the



**Figure 4 | *Saccharomyces* phenotype variation.** A selection of growth phenotypes for *S. cerevisiae* and *S. paradoxus* strains in different environments and drugs. The complete set of lag, rate and density phenotypes in 67 environments is displayed in Supplementary Fig. 9. Phenotypes were quantified using high-resolution micro-cultivation measurements of population density. Strain ( $n = 2$ ) doubling time (rate)

mosaic lineages ( $P = 0.78$ ). Hence, the higher phenotypic variance in *S. cerevisiae* is not driven by out-breeding or domestication per se, but rather suggests that *S. cerevisiae* occupies a wider diversity of ecological niches than *S. paradoxus*.

This survey of *S. cerevisiae* and *S. paradoxus* population genomics reveals extensive differences in genomic and phenotypic variation despite ecological similarities and will allow rapid fine mapping of the genetic determinants. Domestication of *S. cerevisiae* has previously been debated<sup>12</sup>. Our results could be interpreted in two ways. One is as a domestication of one or two groups, the Wine/European and Sake strains, with selection for improved fermentation properties. These domesticated groups then gave rise to feral and clinical derivatives and were involved in the generation of out-crossed derivatives found in all sources. The alternative interpretation is that human activity simply may have used existing strains from populations that had appropriate fermentation properties providing the opportunity to out-breed through movement of strains and supplying a novel disturbed environment. Using domestication to imply 'species bred in captivity'<sup>25</sup>, the strains that best fulfil this definition are the baking isolates, as they have clearly arisen from crosses between lineages. Lineages that were selected from captive bred strains would be expected to have lower diversity than other lineages. This is not the case for the Wine/European or Sake lineages, which have similar or greater levels of diversity in comparison with the other clean lineages or to *S. paradoxus* populations. This view of human activity simply moving yeast strains around without captive breeding is consistent with analysis of over 600 strains<sup>26</sup>. Recent findings in the Malaysian rain-forest (from which our three Malaysian *S. cerevisiae* strains were isolated) of chronic intake of alcoholic nectar from Bertram palms by wild tree shrews suggest that the association of fermented beverages and primates is ancient and not exclusive to humans<sup>27</sup>.

Beyond the analyses we have presented here, the sequence data we have obtained for these strains have many other applications, and have already been used both for global<sup>28</sup> and gene-specific<sup>29</sup> studies. With the advent of new sequencing technology, it is becoming possible to undertake similar population genomic studies for species with much larger genomes, including humans<sup>30</sup>, enabling a new era of genome wide evolutionary and functional genetics.

phenotypes in relation to the S288c derivative BY4741 ( $n = 20$ ) are displayed. Green, poor growth; red, good growth. Hierarchical clustering of phenotypes was performed using a centred Pearson correlation metric and average linkage mapping. Blue, *S. paradoxus*; pink, *S. cerevisiae*; grey, *S. bayanus* isolate CBS7001. LSC, logarithmic strain coefficient.

## METHODS SUMMARY

Strains to be sequenced were selected to maximise the variety of sources and locations of isolation. Except for laboratory strains, a single meiotic diploid spore was isolated from the original strain to remove any heterozygosity<sup>8</sup>. DNA was extracted from overnight cultures<sup>8</sup> for subsequent sequencing on ABI 3730 DNA sequencers and an Illumina Genome Analyzer<sup>15</sup>. Reference-based genome assemblies were created for each strain in a series of steps<sup>15</sup>. Each read was aligned to the reference genome (S288c or CBS432). As this approach cannot deal with large indels or with sequences not present in the reference genome, we developed an iterative parallel-alignment assembling tool, PALAS (Supplementary Methods), to introduce insertions that were allowed to share material between related strains. Two versions of each strain sequence were produced, a partial assembly derived just from data collected from that strain and a more complete assembly using an imputation process to infer the most likely sequence of the strain taking into account data from related strains. In both cases, confidence estimates are given for each base call. The SNPs obtained were used to generate neighbour-joining phylogenetic trees<sup>15</sup>, infer population structure<sup>17</sup>, estimate sequence divergence<sup>10</sup> and analyse polymorphisms<sup>10</sup>. Non-aligned reads (those missing in the reference genome) were searched for potential novel genes. Each strain isolate was subjected to precise phenotyping in 67 experimental conditions using a high-resolution micro-cultivation Bioscreen C (Oy Growth Curves, Finland)<sup>24</sup>. Two consecutive rounds of 48-h pre-cultivation in synthetic complete media were followed by a 72-h cultivation in stress media. Readings of optical density were taken every 20 min. Strains were tested as duplicates ( $N = 2$ ). Growth variables were normalized to the behaviour of the 20 BY4741 replicates.

Details of the methods mentioned above are provided in Supplementary Information.

Received 18 June; accepted 22 December 2008.

Published online 11 February 2009.

- Goffeau, A. et al. Life with 6000 genes. *Science* **274**, 546–567 (1996).
- Mewes, H. W. et al. Overview of the yeast genome. *Nature* **387** (suppl.), 7–8 (1997).
- Sampaio, J. P. & Gonçalves, P. Natural populations of *Saccharomyces kudriavzevii* in Portugal are associated with oak bark and are sympatric with *S. cerevisiae* and *S. paradoxus*. *Appl. Environ. Microbiol.* **74**, 2144–2152 (2008).
- Sniegowski, P. D., Dombrowski, P. G. & Fingerman, E. *Saccharomyces cerevisiae* and *Saccharomyces paradoxus* coexist in a natural woodland site in North America and display different levels of reproductive isolation from European conspecifics. *FEMS Yeast Res.* **1**, 299–306 (2002).
- Aa, E., Townsend, J. P., Adams, R. I., Nielsen, K. M. & Taylor, J. W. Population structure and gene evolution in *Saccharomyces cerevisiae*. *FEMS Yeast Res.* **6**, 702–715 (2006).

6. Koufopanou, V., Hughes, J., Bell, G. & Burt, A. The spatial scale of genetic differentiation in a model organism: the wild yeast *Saccharomyces paradoxus*. *Phil. Trans. R. Soc. Lond. B* **361**, 1941–1946 (2006).
7. Kuehne, H. A., Murphy, H. A., Francis, C. A. & Sniegowski, P. D. Allopatric divergence, secondary contact, and genetic isolation in wild yeast populations. *Curr. Biol.* **17**, 407–411 (2007).
8. Liti, G., Barton, D. B. & Louis, E. J. Sequence diversity, reproductive isolation and species concepts in *Saccharomyces*. *Genetics* **174**, 839–850 (2006).
9. Ruderfer, D. M., Pratt, S. C., Seidel, H. S. & Kruglyak, L. Population genomic analysis of outcrossing and recombination in yeast. *Nature Genet.* **38**, 1077–1081 (2006).
10. Tsai, I. J., Bensasson, D., Burt, A. & Koufopanou, V. Population genomics of the wild yeast *Saccharomyces paradoxus*: Quantifying the life cycle. *Proc. Natl Acad. Sci. USA* **105**, 4957–4962 (2008).
11. Pretorius, I. S. Tailoring wine yeast for the new millennium: novel approaches to the ancient art of winemaking. *Yeast* **16**, 675–729 (2000).
12. Fay, J. C. & Benavides, J. A. Evidence for domesticated and wild populations of *Saccharomyces cerevisiae*. *PLoS Genet.* **1**, 66–71 (2005).
13. Liti, G., Peruffo, A., James, S. A., Roberts, I. N. & Louis, E. J. Inferences of evolutionary relationships from a population survey of LTR-retrotransposons and telomeric-associated sequences in the *Saccharomyces sensu stricto* complex. *Yeast* **22**, 177–192 (2005).
14. Naumov, G. I., James, S. A., Naumova, E. S., Louis, E. J. & Roberts, I. N. Three new species in the *Saccharomyces sensu stricto* complex: *Saccharomyces cariocanus*, *Saccharomyces kudriavzevii* and *Saccharomyces mikatae*. *Int. J. Syst. Evol. Microbiol.* **50**, 1931–1942 (2000).
15. Carter, D. M. *Saccharomyces* genome resequencing project. *Wellcome Trust Sanger Institute* (<http://www.sanger.ac.uk/Teams/Team118/sgrp/>) (2005).
16. Kellis, M., Patterson, N., Endrizzi, M., Birren, B. & Lander, E. S. Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* **423**, 241–254 (2003).
17. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959 (2000).
18. Wei, W. *et al.* Genome sequencing and comparative analysis of *Saccharomyces cerevisiae* strain YJM789. *Proc. Natl Acad. Sci. USA* **104**, 12825–12830 (2007).
19. Schacherer, J., Shapiro, J. A., Ruderfer, D. M. & Kruglyak, L. Comprehensive polymorphism survey elucidates population structure of *Saccharomyces cerevisiae*. *Nature* doi:10.1038/nature07670 (this issue).
20. Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595 (1989).
21. McDonald, J. H. & Kreitman, M. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* **351**, 652–654 (1991).
22. Bulmer, M. The selection-mutation-drift theory of synonymous codon usage. *Genetics* **129**, 897–907 (1991).
23. Clark, R. M. *et al.* Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana*. *Science* **317**, 338–342 (2007).
24. Warringer, J. & Blomberg, A. Automated screening in environmental arrays allows analysis of quantitative phenotypic profiles in *Saccharomyces cerevisiae*. *Yeast* **20**, 53–67 (2003).
25. Diamond, J. Evolution, consequences and future of plant and animal domestication. *Nature* **418**, 700–707 (2002).
26. Legras, J. L., Merdinoglu, D., Cornuet, J. M. & Karst, F. Bread, beer and wine: *Saccharomyces cerevisiae* diversity reflects human history. *Mol. Ecol.* **16**, 2091–2102 (2007).
27. Wiens, F. *et al.* Chronic intake of fermented floral nectar by wild treeshrews. *Proc. Natl Acad. Sci. USA* **105**, 10426–10431 (2008).
28. Mancera, E., Bourgon, R., Brozzi, A., Huber, W. & Steinmetz, L. M. High-resolution mapping of meiotic crossovers and non-crossovers in yeast. *Nature* **454**, 479–485 (2008).
29. Demogines, A., Wong, A., Aquadro, C. & Alani, E. Incompatibilities involving yeast mismatch repair genes: a role for genetic modifiers and implications for disease penetrance and variation in genomic mutation rates. *PLoS Genet.* **4**, e1000103 (2008).
30. Siva, N. 1000 Genomes project. *Nature Biotechnol.* **26**, 256 (2008).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank all the members of the Sanger sequencing production teams for generating the sequence data. We thank members of the Durbin and Louis laboratories and C. Nieduszynski for comments and suggestions, L. Kruglyak and J. Schacherer for sharing their unpublished manuscript and R. Ames and S. Lovell for sharing unpublished results. We also thank the British Council and Chinese Academy of Sciences for providing the opportunity to design and develop this project. Research at the Wellcome Trust Sanger Institute (D.M.C., A.M.M., L.P., M.J., M.A.Q., I.G., S.S., F.S. and R.D.) is supported by The Wellcome Trust. G.L., D.B.H.B., E.B. and E.J.L. were supported by the Wellcome Trust, the Royal Society and the Biotechnology and Biological Sciences Research Council (BBSRC). S.A.J., R.P.D. and I.N.R. were supported by the BBSRC. A.B. and J.W. were supported by the Swedish Research Council and the Swedish Foundation for Strategic Research. A.B. and V.K. were supported by the National Environment Research Council (NERC) and I.J.T. was supported by the Wellcome Trust. D.B. was supported by NERC. A.M.M. was supported by the Canada Foundation for Innovation. M.J.T.O. and A.v.O. were supported by the National Science Foundation, the National Institutes of Health and a Hertz fellowship.

**Author Contributions** R.D. and E.J.L. designed the project. G.L. selected and manipulated yeast strains and extracted DNA samples. M.J., M.A.Q., I.G., S.S. and F.S. performed the subcloning and sequencing. D.M.C. did the reference comparison and assembly of the sequences. D.M.C. and G.L. coordinated the collection of data. D.M.C. and R.D. performed much of the global analysis, which was the basis for specific analyses performed by the other authors. A.M.M. did the selection studies. E.J.L., G.L., D.M.C. R.D., D.B.H.B., E.B. and L.P. did the population structure and analysis of new genes. C.M.B. and D.B. performed the analysis of Ty-element abundance. S.A.J., R.P.D., M.J.T.O., A.v.O. and I.N.R. analysed the rDNA. A.B., V.K. and I.J.T. did the sequence variation and recombination analyses. A.M.M. and A.N.N.B. created a BLAST server. J.W. and A.B. generated the phenomics data. E.J.L. and G.L. wrote the paper, coordinating the contributions of the other authors.

**Author Information** ABI data were submitted to the NCBI Trace Archive (accession numbers: TI1250539685–TI1250559725, TI1253231395–TI1253251114, TI1253289161–TI1253346699, TI1253412335–TI1253476299, TI1253519268–TI1253585710, TI1253601045–TI1253652620, TI1253786059–TI1253860555, TI1253882296–TI1253926729, TI1253930843–TI1253993129, TI1253998006–TI1254046664, TI1254080659–TI1254150525, TI1254155855–TI1254203444, TI1254229054–TI1254316243, TI1254369656–TI1254427407, TI1254441246–TI1254496019, TI1254545851–TI1254582436. The Solexa data were submitted to the European Read Archive (accession number ERA000011). Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to R.D. ([rd@sanger.ac.uk](mailto:rd@sanger.ac.uk)) or E.J.L. ([ed.louis@nottingham.ac.uk](mailto:ed.louis@nottingham.ac.uk)).



## LETTERS

# Comprehensive polymorphism survey elucidates population structure of *Saccharomyces cerevisiae*

Joseph Schacherer<sup>1\*†</sup>, Joshua A. Shapiro<sup>1\*</sup>, Douglas M. Ruderfer<sup>1</sup> & Leonid Kruglyak<sup>1</sup>

Comprehensive identification of polymorphisms among individuals within a species is essential both for studying the genetic basis of phenotypic differences and for elucidating the evolutionary history of the species. Large-scale polymorphism surveys have recently been reported for human<sup>1</sup>, mouse<sup>2</sup> and *Arabidopsis thaliana*<sup>3</sup>. Here we report a nucleotide-level survey of genomic variation in a diverse collection of 63 *Saccharomyces cerevisiae* strains sampled from different ecological niches (beer, bread, vineyards, immunocompromised individuals, various fermentations and nature) and from locations on different continents. We hybridized genomic DNA from each strain to whole-genome tiling microarrays and detected 1.89 million single nucleotide polymorphisms, which were grouped into 101,343 distinct segregating sites. We also identified 3,985 deletion events of length >200 base pairs among the surveyed strains. We analysed the genome-wide patterns of nucleotide polymorphism and deletion variants, and measured the extent of linkage disequilibrium in *S. cerevisiae*. These results and the polymorphism resource we have generated lay the foundation for genome-wide association studies in yeast. We also examined the population structure of *S. cerevisiae*, providing support for multiple domestication events as well as insight into the origins of pathogenic strains.

With their small and compact genomes, the hemiascomycetes (the group of fungi that includes *S. cerevisiae*) represent a powerful model for comparative genomics and studies of genome evolution<sup>4–6</sup>. As a result, more than 18 hemiascomycetes species are either completely or partially sequenced. The availability of the sequence data has presented an unprecedented opportunity to evaluate DNA sequence variation and genome evolution in a phylum spanning a broad evolutionary range<sup>7</sup>. This wealth of data on interspecific sequence differences stands in contrast to our limited knowledge of sequence variation within *S. cerevisiae*. Because of its importance both to human activities and as a model system, we sought to generate a comprehensive view of sequence polymorphism in *S. cerevisiae*. To determine sequence variation at the nucleotide level, we hybridized genomic DNA from 63 ecologically and geographically diverse strains (Supplementary Table 1) to a high-density Affymetrix Yeast Tiling Microarray (YTM) and identified positions likely to differ from the reference sequence with the software package SNPscanner<sup>8</sup>. We detected a total of 1,896,131 single nucleotide polymorphisms (SNPs) in non-repetitive regions of the genome (Supplementary Table 1). Because of variation of up to a few base pairs (bp) in the location of SNPs detected by SNPscanner, we used a grouping procedure (see Methods) to identify the sites of polymorphic variation across strains. We also removed all singletons (SNPs called in only one strain) to reduce false positives further. This approach detected a total of 1,299,811 individual SNP calls, which were grouped into 101,343 distinct segregating sites. At each

of these sites, every strain was classified as having either the same or different nucleotide relative to the reference strain (S288c).

We evaluated the coverage and accuracy of our polymorphism survey by comparing our data with the low-coverage sequence generated by ref. 9; 13 strains are shared between the two data sets. Most of the array-called SNPs with sequence data in the region had corresponding polymorphisms in the sequence data (median of 92% per strain), showing that our data have a low false-positive rate. Array-based polymorphism calls captured most (median of 73% per strain) of the high-quality (quality score >30), independent (>25 bp from the next closest polymorphism) SNPs present in the sequence data, showing that our data have high coverage. Discrepancies between array-based and sequence-based polymorphism calls probably reflect false positives and false negatives in each type of data, and may also derive from genuine sequence differences between strains with the same name but obtained from different sources by the two studies.

We detected an average of 30,097 SNPs per strain (Supplementary Table 1). Excluding laboratory strains, most of which are closely related to the reference strain, the frequency of polymorphisms varied between 0.0011 to 0.0041 per bp (0.0028 on average), representing an average density of 2.8 SNPs per kilobase (kb). Across all strains, we observed 8.35 non-singleton segregating sites per kb ( $\theta_w \text{ kb}^{-1} = 2.26$ ). The frequency spectrum of the observed polymorphisms is highly skewed towards an excess of low-frequency alleles, even after corrections for the grouping procedure and genotyping errors (Supplementary Fig. 1). This excess of rare alleles resulted in a lowered value for the frequency-weighted measure of nucleotide diversity ( $\pi \text{ kb}^{-1} = 1.92$ ). Some of the excess of low-frequency alleles can be attributed to the presence of slightly deleterious variants, which are kept at low frequency by negative selection but have not yet been purged from the population. We expect that mutations in coding regions are more likely to be deleterious than those in noncoding regions, resulting in a lower overall level of polymorphism in coding regions, and we do observe that coding regions are approximately 17% less polymorphic than noncoding regions (Table 1). The coding regions also show a larger excess of low-frequency polymorphism in their frequency spectrum (Supplementary Fig. 2). These trends are further emphasized in the 1,114 genes known to be essential in the reference strain S288c, which show both a lower overall level of polymorphism and a greater skew in the frequency spectrum. Noncoding regions are subject to selection on regulatory elements. Short intergenic regions should carry a higher proportion of functional regulatory sequences than longer noncoding regions, and we observe that intergenic regions shorter than 300 bp have significantly lower rates of polymorphism than longer regions (Supplementary Fig. 3). We found a markedly nonrandom distribution of polymorphism levels across the genome. We observed a decrease in SNP density within 25 kb of centromeres (Supplementary Fig. 4a). This observation is consistent with the lack of

<sup>1</sup>Lewis-Sigler Institute for Integrative Genomics, Department of Ecology and Evolutionary Biology and Howard Hughes Medical Institute, Princeton University, Princeton, New Jersey 08544, USA. <sup>†</sup>Present address: Department of Molecular Genetics, Genomics and Microbiology, Louis-Pasteur University and CNRS, UMR7156, Strasbourg 67083, France.

\*These authors contributed equally to this work.

**Table 1 | Functional variation in levels of polymorphism**

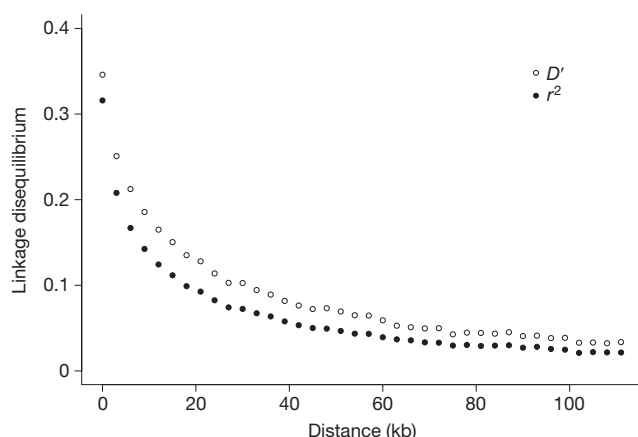
| Genomic region    | Number of segregating sites | $\theta_W$ kb <sup>-1</sup> | $\pi$ kb <sup>-1</sup> |
|-------------------|-----------------------------|-----------------------------|------------------------|
| Noncoding regions | 33,465                      | 2.69                        | 2.33                   |
| Coding regions    | 71,420                      | 2.23                        | 1.89                   |
| Paralogues        | 14,774                      | 2.28                        | 1.93                   |
| Essential genes   | 13,107                      | 2.08                        | 1.77                   |

In a standard neutral model  $4N\mu$  is the population mutation rate, where  $N$  is the effective population size and  $\mu$  the rate of mutation.  $\theta_W$  is Watterson's theta: an estimate of  $4N\mu$  based on the number of segregating sites.  $\pi$  is an estimate of  $4N\mu$  equal to the mean pairwise difference among individuals.

DNA double-strand breaks (that is, the presence of meiotic recombination cold spots) near the centromeres<sup>10</sup>. By contrast, subtelomeric regions, which undergo frequent recombination<sup>11</sup>, show higher variation at the sequence level in the regions 15–45 kb from telomeres (Supplementary Fig. 4b).

The genomic extent of linkage disequilibrium—nonrandom association of alleles at different polymorphic sites—provides information about recombination and population structure, and is also a critical parameter for population studies of association between genotype and phenotype. Our data provided the first opportunity to measure genome-wide properties of linkage disequilibrium across a large collection of diverse strains. We examined pairwise linkage disequilibrium for the 101,343 segregating sites and found that linkage disequilibrium falls to one-half of its maximum value at about 11 kb (Fig. 1). Because the yeast genome is physically compact (12 Mb), the 101,343 segregating sites reported here (nearly a site every 100 bp, of which close to one-half have a minor allele frequency >10%) provide a high-density polymorphism resource for *S. cerevisiae* from which an optimized panel of sites sufficient for whole-genome association studies in yeast can be chosen. To characterize further the architecture of linkage disequilibrium, we examined each of the sampling groups that contained at least 10 strains (wine, clinical, distillery and laboratory strains; Supplementary Fig. 5). In the wine strains, linkage disequilibrium falls to half of its maximum value at ~2.5 kb, but is more extensive in clinical (~7 kb), distillery (~9.5 kb) and laboratory (~23.8 kb) strains. Because most of the laboratory strains are recently derived from the same founder strain S288c<sup>12</sup>, linkage disequilibrium is expected to be greater than in the other groups. By contrast, the low level of linkage disequilibrium in the wine strains probably reflects a long time since the most recent common ancestor of these strains, and perhaps a higher frequency of outcrossing events.

To examine structural variation, we identified all deletion events >200 bp in the 63 strains (Supplementary Tables 1 and 2).

**Figure 1 | Decay of linkage disequilibrium as a function of distance.**

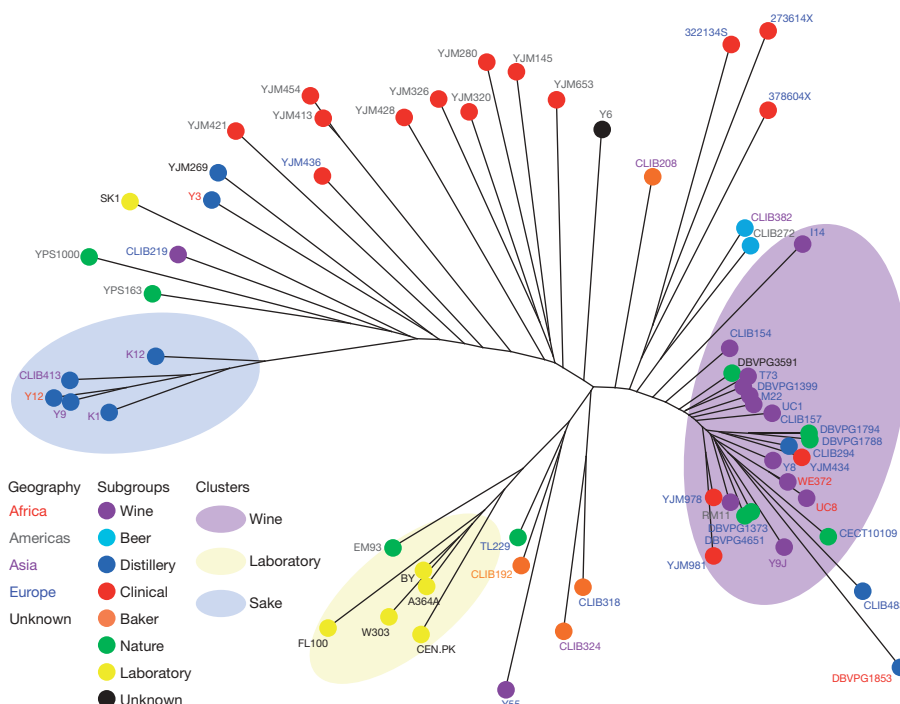
Averages of pairwise linkage disequilibrium measures  $r^2$  (black circles) and  $D'$  (open circles) are plotted for each bin of distances between pairs of SNPs. The linkage disequilibrium values were corrected for finite-size effects by subtracting the average value computed for a random subset of pairs of SNPs located on different chromosomes.

We observed 3,985 deletions (an average of 63 per strain). The number of deletions varied from one in BY4716 (which is isogenic to the reference but carries an engineered deletion of *LYS2*) to 106 in YJM320. The deletions ranged in size from 200 bp to 13.8 kb, with nearly half falling between 200 bp and 400 bp (Supplementary Fig. 6).

The deletions are unevenly distributed across the genome (Supplementary Fig. 7), with enrichment in subtelomeric regions (45.4% of events in <10% of the genome; Supplementary Fig. 7b) and a deficit near the centromeres (Supplementary Fig. 7a). These patterns are consistent with SNP rates and may similarly be explained by variation in recombination rates. A total of 254 genes contained a whole (119 genes) or partial (135 genes) deletion in at least one strain (Supplementary Table 3). Most were deleted in one to four strains, but some were deleted in many strains (Supplementary Fig. 8). For example, the gene YAR047C is deleted in 59 of the 63 surveyed strains. This gene is annotated as a dubious open reading frame (ORF) unlikely to encode a protein, on the basis of comparative sequence data of *Saccharomyces sensu stricto* species<sup>4</sup>. Our observation within the *S. cerevisiae* species strongly supports this hypothesis. Dubious ORFs accounted for 37 of the gene deletions. The set of deleted genes is enriched for those with known functions in transport, and in particular for sugar and hexose transporters (Supplementary Table 4). Most of these deleted genes are located in the subtelomeric regions. These results provide clear evidence of the importance of variation at subtelomeric regions in adaptation of strains to different carbon sources, as previously suggested<sup>12,13</sup>.

We looked for deletions in genes known to be essential in the S288c strains<sup>14</sup>. We observed partial deletions in only four of the 1,114 essential genes (*KRS1*, *PGS1*, *SMT3* and *ERG20*), many fewer than the 49.6 genes that would be expected from the overall deletion frequency ( $\chi^2 = 52$ ;  $P < 0.0001$ ), which shows that the vast majority of the genes defined to be essential in the S288c background are also essential in all other genetic backgrounds of *S. cerevisiae*. With the exception of *KRS1*, these deletions were observed in only a few strains (Supplementary Table 3). Moreover, the deletions observed in the four essential genes affect a small fraction of the ORF, and the genes may still be functional. We examined more closely the partial deletion in *KRS1*, which encodes the lysyl-tRNA synthetase. We looked at the spore viability from crosses between the S288c reference strain and several of the strains (K1, CLIB219, K12 and Y9) in which the *KRS1* gene is partially deleted (Supplementary Fig. 9). We observed a high spore viability of around 90% in each cross, which shows that the *KRS1* gene is still functional in these strains. We also observed a reduced deletion rate in duplicated gene pairs derived from the whole genome duplication event (20 observed compared with 49.4 expected;  $\chi^2 = 21.7$ ;  $P < 0.0001$ ; Supplementary Table 3).

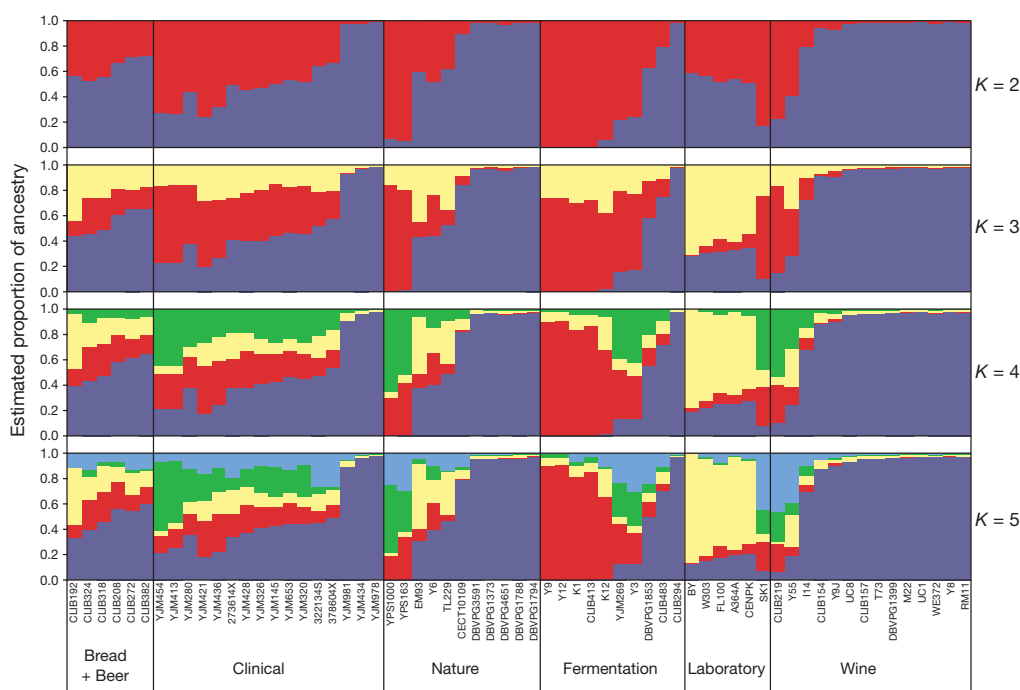
We sought to use the genome-wide genotypes at the 101,343 polymorphic sites across our diverse collection of strains to elucidate the phylogenetic relationships among strains and to evaluate the effects of ecological factors and geographical locations on strain diversity. We used standard neighbour-joining methods to build a majority-rule consensus tree of the surveyed strains (Fig. 2), and also analysed the data with the model-based clustering algorithm implemented in the program Structure<sup>15</sup> (Fig. 3). Both analyses showed at least three distinct subgroups based on the source from which the strains were isolated. Most of the wine strains (with the exception of CLIB219, which was isolated in Russia) are members of a single well-defined subpopulation. Because these wine strains were collected from dispersed locations, this observation provides strong evidence of a single domestication event of yeast for winemaking, followed by human-associated migration of wine yeast all over the world. The wine strains show the lowest level of polymorphism among the groups (Table 2), as well as an excess of low-frequency SNPs, consistent with a bottleneck during domestication. This subpopulation also includes a number of strains collected from distilleries, nature (soil, cocoa beans, prickly pear and *Tuber magnatum*) and clinical sources, indicating that these strains derived from domesticated wine strains, which transited out of



**Figure 2 | Neighbour-joining tree of 63 *S. cerevisiae* strains.** The tree was constructed on the basis of the 101,343 segregating sites identified in the surveyed strains. Branch lengths are proportional to the number of the 101,343 segregating sites that differentiate each pair of strains. Font colour of strain name denotes geographical origin and circle colour denotes ecological niche as specified in the key.

this group to other human-associated fermentations as well as back into nature and therefore escaped their human-manufactured environment. The second major population group contains the strains used for sake production and provides strong evidence for a second and independent domestication event, as hypothesized previously<sup>16</sup>. The laboratory strains, with the exception of SK1, form a third clear group, a consequence of the fact that most of the commonly used *S. cerevisiae* strains, with the exception of SK1, are derived from the S288c genetic background<sup>12</sup>. It is worth noting that the EM93 strain, the progenitor of S288c originally isolated from a rotting fig<sup>17</sup>, is seen to be closely related to the laboratory strains. A number of strains did not fall into clear groups on the tree and did not cluster into coherent groups in the Structure analysis; their genomes seem to be mosaics of contributions from the three genetically distinct subgroups.

Although *S. cerevisiae* is usually considered to be a benign organism, there is a growing recognition that it can be a cause of opportunistic pathogenic fungal infection, typically, but not exclusively, in immunocompromised individuals<sup>18</sup>. To investigate the origin of these strains, we examined 16 strains isolated from different clinical sources (for example, blood, mouth, sputum) in Europe and the Americas (Supplementary Table 1). The clinical isolates were broadly distributed across the tree, and did not cluster with each other or with any one subgroup of strains in the Structure analysis (Figs 2 and 3). Three European clinical strains (YJM434, YJM978 and YJM981) were closely related to wine strains. Three other European strains from the same geographical origin (Newcastle, UK) were closely related to each other, and had some similarity to beer and baker strains. The remaining ten strains (nine American, one European) branched from



**Figure 3 | Population structure of 63 *S. cerevisiae* strains.** Cluster results from a structure analysis on 101,343 segregating sites identified in the 63 surveyed strains. The Structure program implements a Bayesian model-based clustering algorithm that attempts to identify genetically distinct subpopulations on the basis of patterns of allele frequencies<sup>15</sup>. Each strain is represented by a single vertical bar, which is partitioned into *K* coloured segments that represent the strain's estimated ancestry proportion in each of the *K* clusters.



**Table 2 | Polymorphism and SNP frequency within groups**

| Strains    | $\theta_{\text{W}} \text{ kb}^{-1}$ | $\pi \text{ kb}^{-1}$ |
|------------|-------------------------------------|-----------------------|
| All        | 2.26                                | 1.92                  |
| Wine       | 1.38                                | 1.25                  |
| Clinical   | 2.53                                | 2.54                  |
| Distillery | 2.54                                | 2.60                  |
| Nature     | 2.03                                | 1.97                  |

See legend to Table 1.

a similar part of the tree, but did not appear to be closely related to each other or to any other coherent group of strain. Our interpretation of these results is that clinical isolates do not derive from a common ancestor or any one type of strain, but rather represent multiple events in which strains present in the environment opportunistically colonize human tissues. Our data provide strong evidence that wine strains are capable of such colonization, and suggest that strains from other sources (beer, bakery, laboratory, nature) can also do this. These results are consistent with clinical reports of patients infected with *S. cerevisiae* baker's strains and with the strain *Saccharomyces boulardii*, which is used therapeutically to treat diarrhoea and is also sold as a probiotic nutritional supplement<sup>19</sup>. Because the main environmental niches for *S. cerevisiae* in nature are not known, clinical strains might represent the best approximation of the overall species diversity of *S. cerevisiae*.

The polymorphism resource we generated, made freely available in the Yeast SNPs Browser database (<http://gbrowse.princeton.edu/yeastSNP>), enables genome-wide association studies of the phenotypic differences among these and other yeast strains. Phenotypic diversity among yeast isolates is significant, and variation is apparent among the surveyed strains at different levels. The genetic basis of a number of interesting phenotypes can be studied in yeast, including growth at high temperature, sporulation efficiency, telomere length, gene expression and response to drugs<sup>20–24</sup>; these studies can now move from linkage in crosses between two strains to the population level. *S. cerevisiae* provides a powerful model system for studies of complex traits because of the ease with which genetic analyses and phenotyping can be carried out and the ability to engineer and test the effects of individual polymorphisms and their combinations on different genetic backgrounds.

Our analysis also provides insight into the population structure of this yeast species. We show evidence for genetic differentiation of three distinct subgroups based on the source from which the strains were isolated: vineyards, sake and related fermentations, and laboratory strains. Thus, population structure at least partly reflects different ecological niches. Surveys of additional strains are needed to resolve fully the roles of ecology versus geography in the genetic differentiation of this species. Our data strongly support the hypothesis that these three groups represent separate domestication events, and that *S. cerevisiae* as a whole is not domesticated. Finally, our results suggest that *S. cerevisiae* strains from a range of environments are capable of opportunistic colonization of human tissues.

## METHODS SUMMARY

Genomic DNA was extracted from 63 yeast strains (listed in Supplementary Table 1) and hybridized to Affymetrix Yeast Tiling Arrays. We used SNPscanner<sup>8</sup> to identify putative SNPs in each of the 63 strains on the basis of the hybridization intensity at each probe. Because there is error in the precise location of SNP calls made by SNPscanner, we used a grouping procedure (described in Methods) in order to integrate SNP calls across strains and minimize the effects of erroneous positive and negative calls.

We constructed a neighbour-joining tree of the 63 strains from the SNP data using Splitstree<sup>25</sup>, with branch lengths proportional to the number of segregating sites that differentiate each node. To infer the population ancestry of the strains we used Structure<sup>15</sup>, with ancestral population numbers between two and six. We calculated linkage disequilibrium across the genome using two standard metrics:  $D'$  and  $r^2$ , both for the whole genome and for each subpopulation. We calculated

other population genetic summary statistics using code based on the libsequence package<sup>26</sup>, and performed coalescent simulations of genome evolution using FastCoal<sup>27</sup>, with corrections for expected error rates and our grouping procedure.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 18 June; accepted 25 November 2008.**

**Published online 11 February 2009.**

1. The International HapMap Consortium. A haplotype map of the human genome. *Nature* **437**, 1299–1320 (2005).
2. Frazer, K. A. et al. A sequence-based variation map of 8.27 million SNPs in inbred mouse strains. *Nature* **448**, 1050–1053 (2007).
3. Clark, R. M. et al. Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana*. *Science* **317**, 338–342 (2007).
4. Kellis, M., Patterson, N., Endrizzi, M., Birren, B. & Lander, E. S. Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* **423**, 241–254 (2003).
5. Clifften, P. et al. Finding functional features in *Saccharomyces* genomes by phylogenetic footprinting. *Science* **301**, 71–76 (2003).
6. Dujon, B. et al. Genome evolution in yeasts. *Nature* **430**, 35–44 (2004).
7. Dujon, B. Yeasts illustrate the molecular mechanisms of eukaryotic genome evolution. *Trends Genet.* **22**, 375–387 (2006).
8. Gresham, D. et al. Genome-wide detection of polymorphisms at nucleotide resolution with a single DNA microarray. *Science* **311**, 1932–1936 (2006).
9. Carter, D. M. et al. Population genomics of domestic and wild yeasts. *Nature* (submitted).
10. Gerton, J. L. et al. Inaugural article: global mapping of meiotic recombination hotspots and coldspots in the yeast *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* **97**, 11383–11390 (2000).
11. Pryde, F. E., Gorham, H. C. & Louis, E. J. Chromosome ends: all the same under their caps. *Curr. Opin. Genet. Dev.* **7**, 822–828 (1997).
12. Schacherer, J. et al. Genome-wide analysis of nucleotide-level variation in commonly used *Saccharomyces cerevisiae* strains. *PLoS ONE* **2**, e322 (2007).
13. Winzler, E. A. et al. Genetic diversity in yeast assessed with whole-genome oligonucleotide arrays. *Genetics* **163**, 79–89 (2003).
14. Winzler, E. A. et al. Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285**, 901–906 (1999).
15. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959 (2000).
16. Fay, J. C. & Benavides, J. A. Evidence for domesticated and wild populations of *Saccharomyces cerevisiae*. *PLoS Genet.* **1**, 66–71 (2005).
17. Mortimer, R. K. & Johnston, J. R. Genealogy of principal strains of the yeast genetic stock center. *Genetics* **113**, 35–43 (1986).
18. Enache-Angoulvant, A. & Hennequin, C. Invasive *Saccharomyces* infection: a comprehensive review. *Clin. Infect. Dis.* **41**, 1559–1568 (2005).
19. de Llanos, R., Querol, A., Peman, J., Gobernado, M. & Fernandez-Espinar, M. T. Food and probiotic strains from the *Saccharomyces cerevisiae* species as a possible origin of human systemic infections. *Int. J. Food Microbiol.* **110**, 286–290 (2006).
20. Steinmetz, L. M. et al. Dissecting the architecture of a quantitative trait locus in yeast. *Nature* **416**, 326–330 (2002).
21. Deutschbauer, A. M. & Davis, R. W. Quantitative trait loci mapped to single-nucleotide resolution in yeast. *Nature Genet.* **37**, 1333–1340 (2005).
22. Gatbonton, T. et al. Telomere length as a quantitative trait: genome-wide survey and genetic mapping of telomere length-control genes in yeast. *PLoS Genet.* **2**, e35 (2006).
23. Brem, R. B., Yvert, G., Clinton, R. & Kruglyak, L. Genetic dissection of transcriptional regulation in budding yeast. *Science* **296**, 752–755 (2002).
24. Perlstein, E. O., Ruderfer, D. M., Roberts, D. C., Schreiber, S. L. & Kruglyak, L. Genetic basis of individual differences in the response to small-molecule drugs in yeast. *Nature Genet.* **39**, 496–502 (2007).
25. Huson, D. H. & Bryant, D. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **23**, 254–267 (2006).
26. Thornton, K. Libsequence: a C++ class library for evolutionary genetic analysis. *Bioinformatics* **19**, 2325–2327 (2003).
27. Marjoram, P. & Wall, J. D. Fast “coalescent” simulation. *BMC Genet.* **7**, 16 (2006).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We are grateful to all the researchers and institutions, and especially J. Fay, for sharing yeast strains. We thank K. Dolinski and J. Matrese for technical support and D. Gresham for comments on the manuscript. The authors acknowledge discussions with members of the Kruglyak and Botstein laboratories. This work was supported by NIH grant R37 MH059520 and a James S. McDonnell Foundation Centennial Fellowship to L.K., and NIH grant GM071508 to the Lewis-Sigler Institute.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to L.K. ([leonid@genomics.princeton.edu](mailto:leonid@genomics.princeton.edu)).

## METHODS

**Yeast strains.** Yeast strains were obtained from a number of laboratories: J. Fay (Washington University), J. Perez-Ortin (University of Valencia), G. Liti and E. Louis (The University of Nottingham), J. McCusker (Duke University) and Jean-Luc Souciet (Louis-Pasteur University). We also purchased strains from different yeast culture collections: CLIB (Collection de Levures d'Intérêt Biotechnologique), CBS (Centraalbureau voor Schimmelcultures), DBVPG (Dipartimento di Biologia Vegetale e Agroambientale of the University of Perugia) and CECT (Colección Española de Cultivos Tipo). Strains used in this study are listed in Supplementary Table 1.

**SNP identification.** Yeast strains were grown in yeast extract, peptone and dextrose (YPD) medium. Total genomic DNA was purified from 30 ml YPD culture using Qiagen Genomic-Tips 100/G and Genomic DNA buffers as per the manufacturer's instructions. Genomic DNA was digested with DNaseI, labelled and hybridized to Affymetrix Yeast Tiling Arrays (YTMs) as described previously<sup>8</sup>.

We used SNPscanner<sup>8</sup> to identify putative SNPs in each of the 63 strains on the basis of the hybridization intensity of DNA at each probe. SNPs from each strain were independently called against the reference FY3 genome using the following parameters: lod score >2, number of probes covering a base >1, and positive region length >6. These parameters are further described in ref. 8 and in the SNPscanner documentation (<http://genomics-pubs.princeton.edu/SNPscanner/>). With these parameters, we previously showed, using the complete genome sequence of strain YJM789, that 90.1% of true SNPs were detected, with only 49 false-positive SNP calls over the entire genome (a false-positive rate of  $4 \times 10^{-6}$  per bp).

Owing to the 4-bp resolution of the YTMs and the variance associated with DNA hybridization intensities, the SNP position predicted by SNPscanner may fall at varying positions surrounding the actual site of the SNP. This variance required us to perform a grouping procedure, combining all the calls within 6 bp of each other into a single segregating site. As the average density of putative SNPs is 1 per 6.3 bp, the probability of grouping two distinct sites is nontrivial. To reduce this probability, we implemented several heuristic filters: first, to reduce false positives, we required that at least one of the called SNPs in each grouping have a lod score >6. Second, we eliminated possible deletion events by removing putative SNPs with large prediction regions (>100 bp). Finally, we required at least 9 bp between each SNP in a group and the next closest call in the genome. We performed this grouping procedure in a top-down manner, by first grouping the SNPs with the most calls at a given position.

We tested the accuracy of this grouping procedure using a set of known high-confidence SNPs from the completely sequenced genomes of the strains S288c, RM11-1a and YJM145. Specifically, we examined 13,839 SNPs for which YJM145 and RM11-1a had the same allele and differed from the reference sequence. For this set, 12,578 (91%) and 11,518 (83%) SNPs were detected before grouping in YJM145 and RM11-1a, respectively. After grouping, 9,119 SNPs were detected in at least one strain, and 8,086 were correctly called in both strains, from which we infer a false-negative rate of 5.7% per strain, given detection in at least one strain. The grouping procedure almost never separated the same site into multiple sites (one case across the genome), and rarely combined

two distinct sites (394 cases; <5% of sites after grouping). These cases are typically SNPs that are located within 4 bp of each other, closer than the theoretical resolution of the YTMs. We also removed all singletons (SNPs called in only one strain) to reduce false positives further.

**Tree building and Structure analysis.** We constructed a neighbour-joining tree of the 63 strains from the SNP data using the software package Splitstree<sup>25</sup>, with branch lengths proportional to the number of segregating sites that differentiate each node. We ran Structure using the linkage model with the population number parameter,  $K$ , set from 2 to 6, for 100,000 iterations after a burn-in of 100,000 iterations, the first 50,000 of which were run under the free-recombination model<sup>15</sup>.

**Linkage disequilibrium.** We calculated linkage disequilibrium across the genome using two standard metrics:  $D'$  and  $r^2$ . We computed these statistics for all pairs of sites located within a given distance, both for all the strains and within each predefined subpopulation. To correct for finite-size effects and differences in sample size among the subpopulations, we subtracted from each statistic the average value for a random subset of SNP pairs located on different chromosomes (which should not show linkage disequilibrium).

**Polymorphism and divergence statistics.** We calculated population genetic summary statistics of polymorphism using code based on the libsequence package<sup>26</sup>. To correct for the removal of singleton SNPs in the data set, modified estimators of the population mutation parameters  $\theta_W$  and  $\pi$  were used<sup>28</sup>. An analogue of Tajima's  $D$  was calculated as the difference between these modified estimates<sup>29</sup>. To obtain significance values, we simulated under a modified coalescent model as described below, conditioning on the observed number of segregating sites and the approximate length of the sequence. The significance of an observed statistic was then taken to be the probability of observing a more extreme value in at least 10,000 simulations. Divergence rates were calculated from the multiple species alignments of ref. 4. We used PAML<sup>30</sup> to obtain maximum likelihood estimates of the rate of evolution along the *S. cerevisiae* branch after divergence from *Saccharomyces paradoxus*.

Coalescent simulations of genome evolution were performed using FastCoal<sup>27</sup>. Output from each coalescent simulation was run through a series of steps to mirror the sources of error inherent in the SNPscanner data. First, called SNPs were randomly removed with a probability of 5%. The addition of randomly missed calls creates a characteristic dearth of high-frequency SNPs in the data set; simulations under a 5% false-negative rate fit very closely with the observed pattern of polymorphism at high frequency. To correct for incorrectly grouped SNPs, we performed the previously described grouping procedure on all simulated data.

28. Fu, Y. X. Estimating effective population size or mutation rate using the frequencies of mutations of various classes in a sample of DNA sequences. *Genetics* **138**, 1375–1386 (1994).
29. Tajima, F. Evolutionary relationship of DNA sequences in finite populations. *Genetics* **105**, 437–460 (1983).
30. Yang, Z. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**, 555–556 (1997).

## LETTERS

# Antiviral immunity in *Drosophila* requires systemic RNA interference spread

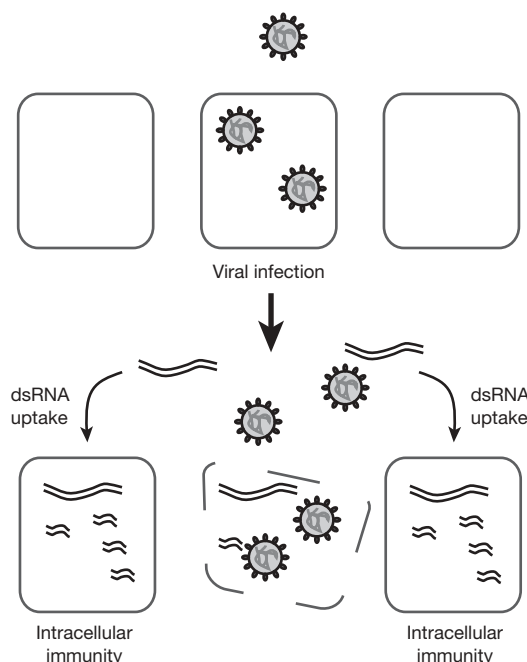
Maria-Carla Saleh<sup>1†</sup>, Michel Tassetto<sup>1\*</sup>, Ronald P. van Rij<sup>1\*†</sup>, Bertsy Goic<sup>2</sup>, Valérie Gausson<sup>2</sup>, Bassam Berry<sup>3</sup>, Caroline Jacquier<sup>3</sup>, Christophe Antoniewski<sup>3</sup> & Raul Andino<sup>1</sup>

Multicellular organisms evolved sophisticated defence systems to confer protection against pathogens. An important characteristic of these immune systems is their ability to act both locally at the site of infection and at distal uninfected locations<sup>1–4</sup>. In insects, such as *Drosophila melanogaster*, RNA interference (RNAi) mediates antiviral immunity<sup>5–7</sup>. However, the antiviral RNAi defence in flies seems to be a local, cell-autonomous process, as flies are thought to be unable to generate a systemic RNAi response<sup>8</sup>. Here we show that a recently defined double-stranded RNA (dsRNA) uptake pathway<sup>9</sup> is essential for effective antiviral RNAi immunity in adult flies. Mutant flies defective in this dsRNA uptake pathway were hypersensitive to infection with *Drosophila* C virus and Sindbis virus. Mortality in dsRNA-uptake-defective flies was accompanied by 100- to 10<sup>5</sup>-fold increases in viral titres and higher levels of viral RNA. Furthermore, inoculating naked dsRNA into flies elicited a sequence-specific antiviral immune response that required an intact dsRNA uptake pathway. These findings suggest that spread of dsRNA to uninfected sites is essential for effective antiviral immunity. Notably, infection with green fluorescent protein (GFP)-tagged Sindbis virus suppressed expression of host-encoded GFP at a distal site. Thus, similar to protein-based immunity in vertebrates, the antiviral RNAi response in flies also relies on the systemic spread of a virus-specific immunity signal.

On the basis of the recent identification of a dsRNA uptake pathway in flies<sup>9,10</sup>, we hypothesized that dsRNA produced and released from infected cells can be taken up locally, and perhaps at distal sites, to establish systemic pre-existing immunity in uninfected cells (Fig. 1). We thus examined whether naked dsRNA can mediate systemic RNAi spread by inoculating flies with dsRNA corresponding to two different regions of the Sindbis virus genome (dsSin1 and dsSin2, Supplementary Fig. 1a; see also Fig. 2a). Two days after dsRNA inoculation, flies were infected with a recombinant Sindbis virus expressing GFP (Sindbis-GFP virus, Supplementary Fig. 1a). Notably, inoculation with dsSin1 and dsSin2 markedly reduced accumulation of GFP as determined by fluorescence microscopy and immunoblotting (Fig. 2b, c, lanes 7–11 and 18–22); control buffer had no effect on virus replication (Fig. 2b, c, lanes 2–6 and Supplementary Fig. 1b). This inhibitory response was sequence specific because flies inoculated with dsRNA corresponding to *Drosophila* C virus (DCV) genome showed no effect on Sindbis virus replication (Fig. 2b, c, lanes 13–17). Furthermore, inoculation of dsRNA corresponding to DCV (dsDCV) efficiently protected wild-type flies against DCV infection, but not against Sindbis virus (Supplementary Fig. 2a). The antiviral effect of exogenous dsRNA inoculation required a functional RNAi machinery as Dicer 2 and Argonaute 2 null mutant flies

(*Dcr2*<sup>−/−</sup> and *Ago2*<sup>−/−</sup>) were unable to mount an effective antiviral response (Fig. 2d and Supplementary Fig. 2a). In addition, wild-type flies accumulated short interfering RNAs (siRNAs) derived from injected dsRNA (Supplementary Fig. 2c). We conclude that inoculation of dsRNA initiates a bona fide, specific RNAi response that protects flies against virus infection.

Serial dilutions of dsSin2 indicated that very low concentrations of injected dsRNA sufficed to mount a very strong response (Fig. 2e). Accordingly, we observed reduced viral replication even after inoculation of 5 pg of dsRNA (equivalent to 1.5 × 10<sup>5</sup> molecules of dsSin2, Fig. 2e, lanes 17–20). Of note, whereas the maximal dose of dsSin2 (5 ng) elicited an inhibitory response that lasted 5 days (Fig. 2e, lanes 5–8), inoculation of a lower dose produced a shorter period of immunity (Fig. 2e, compare lanes 5–8 with 9–12, 13–16 and 17–20).



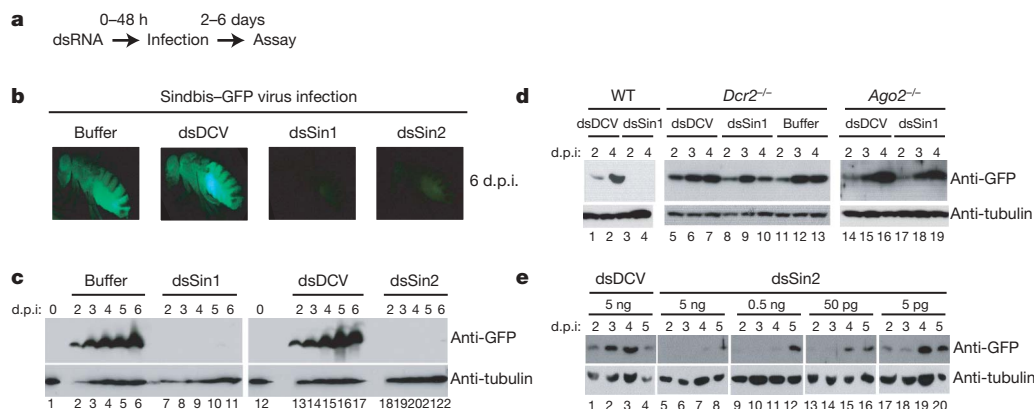
**Figure 1 | Model for systemic RNAi viral immunity in *Drosophila melanogaster*.** Upon viral infection, virus-specific dsRNAs (for example, replication intermediates) are generated during the initial rounds of virus replication. After cell death or lysis, dsRNAs are taken up and processed by uninfected cells to protect them from subsequent infection, thereby preventing virus spread.

<sup>1</sup>Department of Microbiology and Immunology, University of California, San Francisco 94122-2280, USA. <sup>2</sup>Institut Pasteur, Viruses and RNA interference, F-75015 Paris, France.

<sup>3</sup>Institut Pasteur, Drosophila Genetics and Epigenetics; CNRS, URA 2578, F-75015 Paris, France. <sup>†</sup>Present addresses: Institut Pasteur, Viruses and RNA interference, F-75015 Paris, France (M.-C.S.); Department of Medical Microbiology, Nijmegen Center for Molecular Life Sciences, University Medical Center Nijmegen, PO Box 9101, 6500 HB Nijmegen, The Netherlands (R.P.v.R.).

\*These authors contributed equally to this work.





**Figure 2** | *In vivo* dsRNA immunization provides sequence-specific antiviral protection in *D. melanogaster*. **a**, Immunization protocol. **b**, Wild-type flies infected with Sindbis-GFP virus two days after intrathoracic injection of dsRNA against *Drosophila* C virus (DCV, 442 bp long, corresponding to the viral polymerase between nucleotides 5589–6030), dsRNA against Sindbis virus non-structural proteins 1 and 2 (dsSin1, 901 bp long, corresponding to nucleotides 1211–2112) or dsRNA against Sindbis virus corresponding to the non-structural proteins 3 and 4 (dsSin2, 954 bp long, corresponding to nucleotides 5485–6439). Buffer,

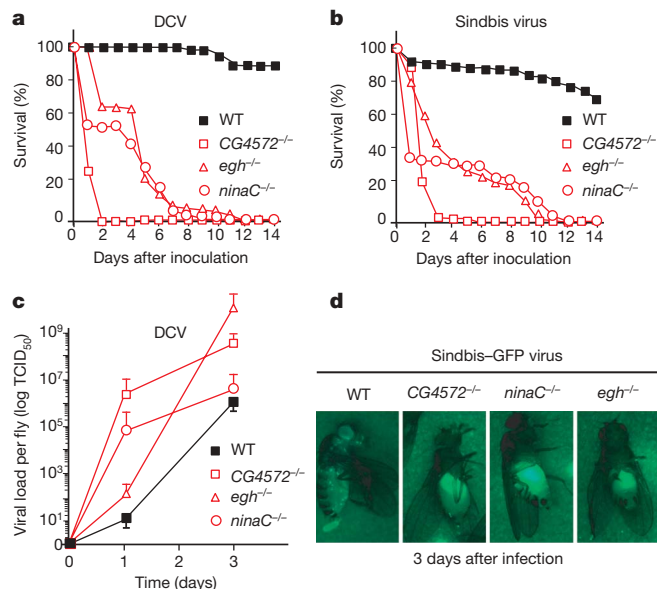
control injection; d.p.i., days post infection. Sindbis-GFP virus replication was monitored by GFP production. **b**, Fluorescence images. **c**, Western blot with an anti-GFP antibody. **d**, Sindbis-GFP virus challenge in wild-type, homozygous *Dcr2*<sup>L811fsX</sup> (*Dcr2*<sup>-/-</sup>) and homozygous *Ago2*<sup>414</sup> (*Ago2*<sup>-/-</sup>) flies. **e**, dsRNA immunization protects in a dose-dependent manner. Flies were inoculated with dsRNA (dsSin2) directed against Sindbis-GFP. Virus replication over time (2–5 d.p.i.) was monitored by western blotting using an anti-GFP antibody.

This observation underscores the efficiency and persistence of the dsRNA-mediated antiviral immunity in *Drosophila*, and supports the idea that exogenous dsRNA can initiate an RNA silencing response in flies, albeit without the RNAi amplification mechanism observed in plants and nematodes<sup>11,12</sup>.

We previously described that dsRNA is taken up in *Drosophila* S2 cells by an active pathway, involving receptor-mediated endocytosis<sup>9</sup>. To examine whether this pathway is involved in the antiviral RNAi response mechanism we selected three genes implicated by the previous analysis in dsRNA uptake: *egghead* (*egh*), encoding a seven-transmembrane-domain glycosyltransferase; *ninaC*, coding for a protein involved in vesicle transport; and a gene of unknown function, *CG4572* (Supplementary Fig. 2b). Although viability and fertility of homozygous *egh*<sup>EP804</sup>, *ninaC*<sup>3</sup> and *CG4572*<sup>c05963</sup> mutant flies did not differ significantly from wild type, all three mutants were hypersensitive to DCV or Sindbis virus infection. In these dsRNA uptake-defective flies, we observed an earlier onset of disease (Fig. 3a, b). After infection, median survival of homozygous *egh*<sup>EP804</sup>, *ninaC*<sup>3</sup> and *CG4572*<sup>c05963</sup> flies was approximately 5–8 days, compared with more than 14 days in wild-type flies, and the 50% lethal dose (LD<sub>50</sub>) in *CG4572*<sup>c05963</sup> flies was ninefold lower than in wild-type flies (not shown).

An important consideration when studying viral sensitivity in animals defective for components of a major cellular pathway, such as endocytosis or intracellular transport, is that enhanced death after viral infection may be caused by a decrease in fitness or general health of the mutant animal, and not by a direct antiviral activity of the deleted component. To establish whether the increased mortality of *egh*, *ninaC* and *CG4572* mutant flies stems from their inability to control virus replication, we determined viral loads (Fig. 3c). Even at early time points after infection, before the onset of disease, DCV titres were 100- to 10<sup>5</sup>-fold higher in homozygous *egh*<sup>EP804</sup>, *ninaC*<sup>3</sup> and *CG4572*<sup>c05963</sup> flies compared to wild-type controls (Fig. 3c). The increase in viral titres in mutant flies was mirrored by a marked increase in viral RNA levels. Whereas viral RNA was barely detected in wild-type flies before day 5, it was clearly observed at 24 h after infection in homozygous *ninaC*<sup>3</sup> and *CG4572*<sup>c05963</sup> mutant flies, and by 48 h it accumulated at much higher levels in these mutants than in wild-type flies (Supplementary Fig. 3a). We examined the role of the dsRNA uptake pathway on virus replication further by monitoring Sindbis-GFP virus tissue tropism. In wild-type flies, GFP fluorescence was barely detected 3 days after infection and accumulated

in discrete puncta throughout the fly. In contrast, in homozygous *egh*<sup>EP804</sup>, *ninaC*<sup>3</sup> and *CG4572*<sup>c05963</sup> mutants, GFP accumulated within a large structure in the abdomen of the animal and at much higher levels than in wild-type flies (Fig. 3d and Supplementary Fig. 3b). These results indicate that the enhanced viral susceptibility of *egh*, *ninaC* and *CG4572* mutant flies is due to their inability to control virus replication.



**Figure 3** | Increased viral susceptibility of dsRNA-uptake-deficient mutants. **a**, **b**, Survival of dsRNA uptake mutant flies after virus infection. Homozygous *egh*<sup>EP804</sup> (*egh*<sup>-/-</sup>), *ninaC*<sup>3</sup> (*ninaC*<sup>-/-</sup>), *CG4572*<sup>c05963</sup> (*CG4572*<sup>-/-</sup>) and wild-type flies were injected with 500 TCID<sub>50</sub> (50% tissue culture infectious dose) DCV (**a**) or 500 plaque-forming units Sindbis-GFP virus (**b**) and monitored daily for survival. **c**, DCV replicates at higher levels in dsRNA uptake mutant flies. Flies were injected with 500 TCID<sub>50</sub> DCV, and virus production was monitored over time. At each time point, three pools of five flies were homogenized, and the viral titre in the homogenate was determined by end-point dilution. The error bars report the average  $\pm$  s.d. for at least three independent experiments. **d**, Sindbis-GFP virus replicates at higher levels in dsRNA uptake mutant flies as shown by increased GFP expression in the fat body at 3 days after infection when compared with wild-type flies.

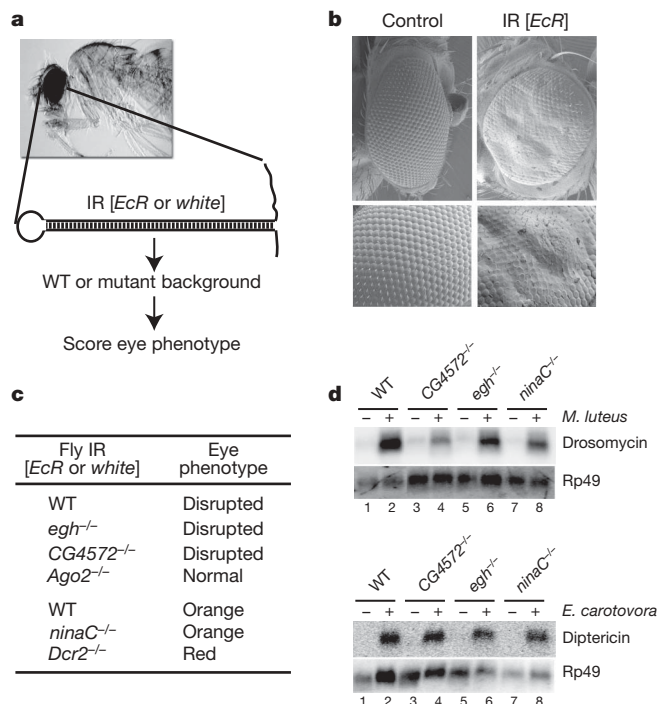
The hyper-susceptibility to virus infection of flies defective in dsRNA uptake is strikingly similar to that previously seen in *Ago2*-defective flies<sup>6</sup>. Therefore, we examined whether the RNAi core function is impaired in *egh*, *ninaC* and *CG4572* mutant flies. Eye-specific silencing of the *Ecdysone receptor* gene (*EcR*) by an endogenously expressed *EcR* hairpin dsRNA<sup>13</sup> leads to abnormal eye structure resulting from impaired corneal lens formation (Fig. 4a, b). Under these conditions, disruption of the core RNAi machinery in homozygous *Ago2*<sup>414</sup> mutant flies suppressed *EcR* RNAi and restored normal eye structure. In contrast, efficient *EcR* RNAi was observed in homozygous *egh*<sup>EP804</sup> and *CG4572*<sup>c05963</sup> flies. Similar experiments monitored RNAi in homozygous *ninaC*<sup>3</sup> flies using the expression of a hairpin dsRNA targeting the *white* gene that causes a decrease of eye pigmentation and orange eye colour in control flies (Supplementary Fig. 4a, b)<sup>14</sup>. Silencing of *white* was suppressed in homozygous *Dcr2*<sup>L811f5X</sup> mutant flies whereas it was fully maintained in homozygous *ninaC*<sup>3</sup> flies (Fig. 4c and Supplementary Fig. 4a, b). We further confirmed this conclusion by injecting dsRNA against the *fushi tarazu* gene (*ftz*)<sup>15</sup> into syncytial embryos before cellularization. Injection of *ftz* dsRNA in wild-type embryos resulted in the expected segmentation defects, namely loss of denticle belts in the cuticle of pre-hatching larvae (*ftz* phenotype; Supplementary Fig. 4c). Injection of *ftz* dsRNA in homozygous *egh*<sup>EP804</sup>, *ninaC*<sup>3</sup> and *CG4572*<sup>c05963</sup> embryos induced the same defects, indicating that RNA silencing proceeded normally in these mutants. In contrast, homozygous *Ago2*<sup>414</sup> control embryos were unable to silence *ftz* expression and thus hatched with a wild-type cuticle (Supplementary Fig. 4c). These results indicate that mutant flies support efficient RNAi silencing if dsRNA uptake is bypassed through expression of dsRNA hairpins intracellularly or by injecting dsRNA into syncytial embryos.

We next examined whether other arms of the immune system were affected in *egh*, *ninaC* and *CG4572* mutants. Insects produce a number of antimicrobial peptides, which are secreted into the haemolymph, in response to immune challenge. These peptides are effective against Gram-negative and Gram-positive bacteria as well as fungi<sup>16</sup>. We determined whether *egh*, *ninaC* and *CG4572* mutant flies can support production of the antimicrobial peptides drosomycin and dipterin in response to septic injury with Gram-positive and Gram-negative bacteria. Drosomycin production was measured after septic injury with *Micrococcus luteus*, a Gram-positive bacterium that signals through the Toll pathway. Production of dipterin was measured after septic injury with *Erwinia carotovora* (also called *Pectobacterium carotovorum*), a Gram-negative bacterium that induces the Imd pathway. Homozygous *egh*<sup>EP804</sup>, *ninaC*<sup>3</sup> and *CG4572*<sup>c05963</sup> flies were able to respond efficiently to bacterial infection, (Fig. 4d). Similarly, the JAK/STAT signalling pathway<sup>17</sup> seems to be unimpaired in *egh*<sup>EP804</sup> and *CG4572*<sup>c05963</sup> flies as DCV infection induced normal *vir-1* expression in these mutants (Supplementary Fig. 5). Thus, defects in cellular components that abrogate dsRNA uptake and its ensuing antiviral immunity do not generally impair other arms of the fly innate immune system.

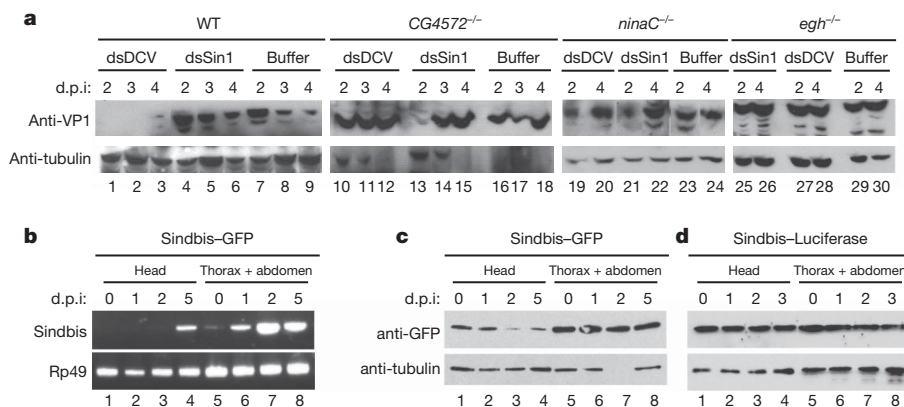
We hypothesize that the dsRNA uptake pathway underlies systemic antiviral immunity, which is required to control virus replication. We thus examined whether dsRNA inoculation in *egh*, *ninaC* and *CG4572* mutant flies was able to elicit the protective immunity observed in wild-type flies (Fig. 2). Indeed, whereas inoculation of DCV dsRNA markedly reduced DCV replication in wild-type flies (Fig. 5a, lanes 1–9), homozygous *egh*<sup>EP804</sup>, *ninaC*<sup>3</sup> and *CG4572*<sup>c05963</sup> mutant flies were unable to mount an antiviral response on DCV dsRNA inoculation (Fig. 5a, lanes 10–30). Similarly, the dsRNA uptake pathway was required for protection against Sindbis virus infection by naked dsSin1 inoculation (Supplementary Fig. 6). Furthermore, whereas wild-type flies efficiently processed inoculated dsRNA into siRNAs, *egh*<sup>EP804</sup>, *ninaC*<sup>3</sup> and *CG4572*<sup>c05963</sup> mutant flies accumulated siRNAs at much lower levels (Supplementary Fig. 2b).

Our model states that infected cells release viral dsRNA that is subsequently taken up by uninfected cells through the dsRNA uptake pathway, thereby eliciting an antiviral RNAi response. A direct prediction of this model is that during infection, viral-derived dsRNA spreads to induce systemic silencing. To test this prediction we examined whether infection with a Sindbis virus carrying the GFP gene could silence a ubiquitously expressed endogenous GFP at a distal site. After intrathoracic inoculation, Sindbis-GFP virus RNA was readily detected in the thorax and abdomen of Tub-eGFP transgenic flies starting at 1 day after infection (Fig. 5b, lanes 5–8). In contrast, the viral RNA was not detectable in the head until day 5 after infection (Fig. 5b, lanes 1–4). Notably, endogenous GFP expression in the head was significantly reduced already at day 2, despite the absence of any detectable viral replication in this organ (Fig. 5c, lane 3). In contrast, infection with control Sindbis virus carrying a firefly luciferase gene did not silence GFP expression (Fig. 5d). These results indicate that a virus-specific derived RNAi signal spreads from the thorax to the head early after infection.

It was previously thought that *Drosophila* is unable to spread systemically an RNAi response, based on observations that endogenously expressed RNA hairpins do not spread from cell to cell<sup>8</sup>. However, we demonstrate that, upon virus infection, infected cells spread systemically a silencing signal that elicits protective RNAi-dependent immunity throughout the organism. Although uninfected *Drosophila* cells seem to lack a constitutive mechanism for systemic RNAi spread, unlike plant and nematode cells, they do have an active and highly efficient mechanism for dsRNA uptake, which we here show is essential for antiviral immunity. Accordingly, dsRNA is normally not released from uninfected cells, but virus infection may induce dsRNA release either through lysis of infected cells or through a virally induced shedding mechanism. We propose that these virally derived dsRNAs are taken up into uninfected cells to generate virus-specific intracellular



**Figure 4 | Core RNAi machinery and antibacterial immunity are intact in dsRNA uptake mutants.** **a**, Schematic to test the core RNAi machinery integrity. **b**, RNAi processing of an inverted repeat IR [*EcR*] induced by the GMR-GAL4 driver prevents the formation of the corneal lens (electron microscopy picture, top panels are  $\times 200$  original magnification, bottom panels are  $\times 500$  original magnification). **c**, Monitoring corneal lens formation and eye colour in transgenic flies deficient in the dsRNA uptake pathway. **d**, Susceptibility of dsRNA uptake mutant flies to infection is specific to the viruses, as the dsRNA uptake mutant flies are able to produce antimicrobial peptides in response to an infection by pathogenic Gram-positive and Gram-negative bacteria.



**Figure 5 | Systemic spread of dsRNA follows virus infection and is essential for effective antiviral immunity.** **a**, *Drosophila* C virus infection in wild-type flies, and in homozygous *egh<sup>EP804</sup>* (*egh<sup>-/-</sup>*), *ninaC<sup>3</sup>* (*ninaC<sup>-/-</sup>*) and *CG4572<sup>c05963</sup>* (*CG4572<sup>-/-</sup>*) mutant flies inoculated with the indicated dsRNA. DCV replication was monitored by western blotting using an antibody directed against DCV capsid protein VP1. **b–d**, dsRNA produced during virus replication can spread and silence endogenous GFP expressed

at a distal site of infection. Flies expressing eGFP (Tub-eGFP) inoculated with Sindbis-GFP (**b**, **c**) or Sindbis-luciferase virus (**d**) by intrathoracic inoculation. **b**, Viral replication monitored by RT-PCR using primers that amplify NSP1/2 virus genes. **c**, Expression of endogenous GFP was monitored by western blot with an anti-GFP antibody. **d**, Same as **c** except that flies were infected with Sindbis-luciferase virus.

immunity that prevents virus spread (Fig. 1). In support of this idea, this specific antiviral response in flies requires both the RNAi core machinery and the recently described dsRNA uptake pathway. Furthermore, simple inoculation of even very low amounts of dsRNA, in the absence of virus infection, can by itself promote a potent antiviral immunity, which is similarly dependent on the RNAi core machinery and the dsRNA uptake pathway. Our previous results indicated that whereas dsRNA is readily taken up by *Drosophila* S2 cells, siRNAs are not efficiently taken up<sup>9</sup>. We thus conclude that systemic spread of a specific antiviral RNAi activity—probably mediated by large viral dsRNAs or intramolecular base-pairing structures released from infected cells—is an essential component of the immune response elicited by virus infection in flies. The precise nature of the RNAi spread intermediate remains to be defined further.

It is remarkable that blocking the spread of the RNAi signal has such a profound effect on antiviral immunity. This suggests that the cell-autonomous RNAi response is insufficient to control a viral infection. In striking parallel to vertebrates, flies also rely on systemic immunity, albeit in this case the virus-specific signal is dsRNA-based. These observations provide an insight into the evolutionarily conserved principles of immunity in multicellular organisms, requiring both cell-autonomous responses as well as systemic mechanisms to create pre-existing immunity to protect uninfected cells.

## METHODS SUMMARY

**Fly stocks.** dsRNA uptake mutant stocks were obtained from the Bloomington *Drosophila* Stock Center. The genomic structure of these mutant allele stocks was confirmed by inverse PCR and sequencing. The *egh<sup>EP804</sup>* allele is a P-element insertion in the coding sequence of the *egh* first exon. The *ninaC<sup>3</sup>* allele is a replacement of the K1078 codon by a stop codon<sup>18</sup>. The *CG4572<sup>c05963</sup>* allele is a PiggyBac insertion in the open reading frame of *CG4572<sup>c05963</sup>*. The UAS>IR[*Ecr*] transgene producing *Ecr* dsRNA<sup>13</sup> and the P[GAL4-ninaE.GMR]12 GAL4 driver were recombined on chromosome 2 before genetic crosses with *egh<sup>EP804</sup>* and *CG4572<sup>c05963</sup>* mutant stocks. The GMR>IR[white] inverted repeat transgene has been previously described<sup>14</sup>. The Tub-eGFP transgenic line was obtained from S. Cohen<sup>19</sup>.

**dsRNA preparation and injection into adult flies.** dsRNA was generated by *in vitro* transcription using T7 RNA polymerase. Five-day-old female flies were injected in the thorax using a nanoinjector (Nanoject II, Drummond Scientific). Two days later flies were infected in the opposite side of the thorax with the appropriate virus.

**Viruses.** Recombinant Sindbis virus expressing GFP during viral replication was generated by cloning enhanced GFP into the XbaI site of the double subgenomic Sindbis vector pTE3'2J (provided by C. Rice)<sup>20</sup>. *In vitro* transcribed RNA was transfected into BHK-21 cells. Virus titre was determined by plaque assay on BHK cells. DCV stock preparation and viral titre calculation have been previously described<sup>6</sup>.

**Microbial infection.** Log phase of growth cultures of bacteria *Erwinia carotovora* and *Micrococcus luteus* were re-suspended in culture medium, and sharpened needles dipped into these suspensions. Flies were harvested at 6 and 36 h after septic injury. RNA extraction and northern blots were performed following standard procedures.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 4 October; accepted 3 December 2008.

Published online 8 February 2009.

- Baulcombe, D. RNA silencing in plants. *Nature* **431**, 356–363 (2004).
- Dörner, T. & Radbruch, A. Antibodies and B cell memory in viral immunity. *Immunity* **27**, 384–392 (2007).
- Roitt, I., Brostoff, J. & Male, D. *Immunology* (Mosby, 2001).
- Voinnet, O. Non-cell autonomous RNA silencing. *FEBS Lett.* **579**, 5858–5871 (2005).
- Galiana-Arnoux, D., Dostert, C., Schneemann, A., Hoffmann, J. A. & Imler, J. L. Essential function *in vivo* for Dicer-2 in host defense against RNA viruses in *Drosophila*. *Nature Immunol.* **7**, 590–597 (2006).
- van Rij, R. P. *et al.* The RNA silencing endonuclease Argonaute 2 mediates specific antiviral immunity in *Drosophila melanogaster*. *Genes Dev.* **20**, 2985–2995 (2006).
- Wang, X. H. *et al.* RNA interference directs innate immunity against viruses in adult *Drosophila*. *Science* **312**, 452–454 (2006).
- Roignant, J. Y. *et al.* Absence of transitive and systemic pathways allows cell-specific and isoform-specific RNAi in *Drosophila*. *RNA* **9**, 299–308 (2003).
- Saleh, M. C. *et al.* The endocytic pathway mediates cell entry of dsRNA to induce RNAi silencing. *Nature Cell Biol.* **8**, 793–802 (2006).
- Ulvila, J. *et al.* Double-stranded RNA is internalized by scavenger receptor-mediated endocytosis in *Drosophila* S2 cells. *J. Biol. Chem.* **281**, 14370–14375 (2006).
- Vaistij, F. E., Jones, L. & Baulcombe, D. C. Spreading of RNA targeting and DNA methylation in RNA silencing requires transcription of the target gene and a putative RNA-dependent RNA polymerase. *Plant Cell* **14**, 857–867 (2002).
- Sijen, T. *et al.* On the role of RNA amplification in dsRNA-triggered gene silencing. *Cell* **107**, 465–476 (2001).
- Schubiger, M., Carre, C., Antoniewski, C. & Truman, J. W. Ligand-dependent depression via *Ecr*/USP acts as a gate to coordinate the differentiation of sensory neurons in the *Drosophila* wing. *Development* **132**, 5239–5248 (2005).
- Lee, Y. S. *et al.* Distinct roles for *Drosophila* Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell* **117**, 69–81 (2004).
- Kennerdell, J. R. & Carthew, R. W. Use of dsRNA-mediated genetic interference to demonstrate that *frizzled* and *frizzled 2* act in the wingless pathway. *Cell* **95**, 1017–1026 (1998).
- Hoffmann, J. A. The immune response of *Drosophila*. *Nature* **426**, 33–38 (2003).
- Dostert, C. *et al.* The Jak-STAT signaling pathway is required but not sufficient for the antiviral response of *Drosophila*. *Nature Immunol.* **6**, 946–953 (2005).
- Porter, J. A., Minke, B. & Montell, C. Calmodulin binding to *Drosophila* NinaC required for termination of phototransduction. *EMBO J.* **14**, 4450–4459 (1995).
- Brennecke, J., Hipfner, D. R., Stark, A., Russell, R. B. & Cohen, S. M. bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene *hid* in *Drosophila*. *Cell* **113**, 25–36 (2003).



20. Hahn, C. S., Hahn, Y. S., Braciale, T. J. & Rice, C. M. Infectious Sindbis virus transient expression vectors for studying antigen processing and presentation. *Proc. Natl Acad. Sci. USA* **89**, 2679–2683 (1992).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We are grateful to members of the Andino and the O'Farrell laboratories for discussion, technical support and advice on fly work. We thank J. Frydman, P. O'Farrell and M. Vignuzzi for discussions and comments on the manuscript. We also thank T. Cook for advice on the IR *EcR* eye phenotype, R. Carthew for providing the GMR>IR[white] and *Dcr2* fly stocks and M. Siomi for the Ago2 fly stock. B.G. is a Manlio Cantarini fellow. B.B. is a Lebanese CNRSL Fellow. C.J. is a University of Paris VI and Ministère de la Recherche fellow. This

work was financially supported by NIH grants AI40085 and AI064738 to R.A., the Institut Pasteur to M.-C.S. and C.A., and CNRS, ANR and ARC grants to C.A.

**Author Contributions** M.-C.S., M.T. and R.P.v.R. performed dsRNA inoculations and virus infections, normal and reverse northern blotting, western blotting, survival curves, obtained fluorescent images, and prepared and analysed mutant flies. B.G. and V.G. examined systemic spread of dsRNA. The genetic and phenotypic analyses of transgenic flies expressing RNA hairpins were designed and carried out by B.B., C.J. and C.A. M.-C.S. and R.A. designed the experiments, discussed the interpretation of the results and co-wrote the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to R.A. ([raul.andino@ucsf.edu](mailto:raul.andino@ucsf.edu)).

## METHODS

**Cells, plasmids and viruses.** *Drosophila* S2 cells (Invitrogen) were cultured at 25 °C in Schneider's *Drosophila* medium (Gibco), supplemented with 10% heat-inactivated fetal calf serum, 2 mM L-glutamine, 100 U ml<sup>-1</sup> penicillin, and 100 mg ml<sup>-1</sup> streptomycin. Firefly (*Photinus pyralis*) and *Renilla reniformis* luciferase sequences from the plasmids pGL3 and pRL-CMV (Promega) were cloned into pMT/V5-HisB (Invitrogen), generating pMT-Luc and pMT-Ren allowing copper-inducible expression from a metallothionein promoter.

Transfections were performed using Effectene transfection reagent (Qiagen) according to the manufacturer's recommendations. Luciferase expression was assayed using the Dual-Luciferase Reporter Assay System (Promega) and analysed on a Tecan Ultra-evolution plate reader. Double-stranded RNA was generated by *in vitro* transcription from T7-promoter-flanked PCR products. DCV viral stocks were prepared on low-passage S2 cells and titred by end-point dilution. Briefly, 25,000 S2 cells per well in a 96-well plate were inoculated with tenfold dilutions of viral stocks. Cells were transferred to fresh medium at day 7 and cytopathic effect was monitored visually over 14 days. Viral titres were calculated according to the method of Reed and Muench<sup>21</sup>.

Recombinant Sindbis virus expressing GFP during viral replication was generated by cloning enhanced GFP into the XbaI site of the double subgenomic Sindbis vector pTE3'2J (provided by C. Rice)<sup>20</sup>. The resulting plasmid was linearized and *in vitro* transcribed using the mMessage machine kit (Ambion). RNA was purified and electroporated into BHK-21 cells, and supernatant was harvested and virus titre determined by plaque assay on BHK cells.

**RNAi in S2 cells.** The effect of downregulating NinaC, CG4572 and Egghead on dsRNA uptake was analysed in a silencing of luciferase expression assay. Cells were pre-treated with approximately 500-nucleotide-long dsRNA targeting *egh* (nucleotides 488–1103; 616-bp product), *ninaC* (nucleotides 161–761; 601-bp product), *cg4572* (nucleotides 61–731; 671-bp product), or *Ago2* (nucleotides 214–865; 652-bp product), or with dsRNA targeting GFP as a negative control. Three days after knockdown of these gene products, the cells were co-transfected with an RNAi dual reporter system, consisting of firefly luciferase and *Renilla* luciferase expression plasmids. Then, dsRNA directed against firefly luciferase (nucleotides 66–658; 592-bp product) was either added to the culture supernatant (soaking) or directly introduced into cells by co-transfection with the dual reporter plasmids (transfection). Twenty-four hours after dsRNA luciferase treatment, expression of luciferase was induced by adding CuSO<sub>4</sub> to the culture supernatant, and cell lysates were generated after an additional 18 h incubation.

**Microbial infection.** The bacteria *Erwinia carotovora* and *Micrococcus luteus* were pre-cultured in LB medium. Pellets taken when the cultures were in the log phase of growth were re-suspended in a small amount of culture medium, and sharpened needles dipped into these suspensions. Flies were harvested at 6 and 36 h after septic injury. Total RNA extraction and northern blots were performed following standard procedures.

**Fly infections.** Flies were reared on standard medium at 25 °C. *Ago2*<sup>414</sup> and *Dcr2*<sup>L811f5X</sup> flies have been described previously<sup>22</sup>. *w*<sup>1118</sup> flies were used as wild-type controls. Two-to-three-day-old female flies were injected with 50 nl of the appropriate virus dilution in 10 mM Tris-Cl (pH 7.5) as described previously<sup>23</sup>, using a Drummond nanoject injector. Fly mortality at day one was attributed to damage invoked by the injection procedure, and these flies were excluded from further analyses. Mortality was monitored daily for 14 days, and every three to four days the flies were transferred to fresh food. In all experiments 40–60 flies per genotypic group were injected. Unless noted otherwise, female flies were used. No significant difference in survival was observed between flies after injection of buffer (data not shown). For northern blots, RNA was isolated from 25 flies using

Trizol reagent. Viral titres in the flies were determined by end-point dilution of fly homogenate of three pools of five flies. At the indicated time points, flies were harvested and stored at -70 °C until further processing. We confirmed the absence of endogenous virus in fly stocks by titration of uninfected fly homogenate on S2 cells.

**dsRNA preparation and injection into adult flies.** dsRNA was generated by *in vitro* transcription using T7 RNA polymerase using as template PCR products corresponding to nucleotides 1211–2112 (NSP1/2) and 5485–6439 (NSP3/4) of the Sindbis virus genome, or nucleotides 5589–6030 of the DCV genome. Five-day-old female flies were CO<sub>2</sub>-anaesthetized and injected in the thorax with 50 nl of the appropriate dsRNA (5 mg ml<sup>-1</sup>) using a nanoinjector (Nanoject II, Drummond Scientific). Two days later flies were CO<sub>2</sub>-anaesthetized and injected in the opposite side of the thorax with the appropriate virus dilution in 10 mM Tris-Cl (pH 7.5). Injection of the same volume of 10 mM Tris-HCl, pH 7.5, was used as a control. Age of the flies and amount of dsRNA injected was determined according to ref. 24. Virus infection has been described previously<sup>6</sup>.

**Western blot analysis.** For protein analysis, equal amounts of protein from total fly extracts were boiled in Laemmli buffer and loaded on 10% SDS-PAGE. After transfer nitrocellulose membranes were blocked in 5% milk, 1× PBS, 0.1% Tween, and incubated overnight with rabbit polyclonal anti-GFP (Santa Cruz Biotechnology) or rabbit polyclonal anti-VP1 (custom made). For normalization a monoclonal antibody anti- $\alpha$ -tubulin (Sigma Aldrich) was used. Detection was performed using Supersignal West Pico Chemiluminescent Substrate (Pierce).

**Northern blots.** Total RNAs were extracted from whole flies using Trizol (Invitrogen). 15  $\mu$ g of total RNA was size fractionated on 1% (w/v) agarose gels containing 1.1 mM formaldehyde. After electrophoresis, the RNA was transferred overnight by capillarity to a nylon membrane (Nytran Supercharge; Schleicher and Schuell) and covalently bound to the membrane using a Stratalinker UV crosslinker. Northern blots were hybridized with DNA probes generated by a random-primed labelling reaction and [ $\alpha$ -<sup>32</sup>P]dCTP. Membranes were exposed overnight to a PhosphorImager screen at room temperature. Viral RNA was detected by northern blot using standard procedures with a random primed DNA probe corresponding to nucleotides 1947–2528 of DCV.

**Oligonucleotide primers.** All the primers used to produce dsRNA had a T7 promoter sequence (TAATACGACTCACTATAGGGAGA) at the 5' end. DCVpol forward, 5'-CAACGAATATGTCGCCTTGA-3'; DCVpol reverse, 5'-TTGGTTGTACGTCAAATCTGAG-3'; SINsp1 forward, 5'-TCTGCCGATCATAGCACAAG-3'; SINsp2 reverse, 5'-CCTTCTTAACGCAACGCTTC-3'; SINsp3 forward, 5'-GAGGATCAATTTTCGACGGAGA-3'; SINsp4 reverse, 5'-GATTGAATGTCGCTGAGTCCAG-3'; *vir-1* forward, 5'-TTCGATTCCTCAGACGATGA-3'; *vir-1* reverse, 5'-GGTCAATGGGCACAAAGTTC-3'; *Rp49* forward, 5'-AAGGGTATCGACAACAGAGTGC-3'; *Rp49* reverse, 5'-ACAAATGTGTATTCCGACCACG-3'.

- Reed, L. J. & Muench, H. A simple method of estimating fifty percent endpoints. *Am. J. Hyg.* **27**, 493–497 (1938).
- Okamura, K., Ishizuka, A., Siomi, H. & Siomi, M. C. Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. *Genes Dev.* **18**, 1655–1666 (2004).
- Cherry, S. & Perrimon, N. Entry is a rate-limiting step for viral infection in a *Drosophila melanogaster* model of pathogenesis. *Nature Immunol.* **5**, 81–87 (2004).
- Goto, A., Blandin, S., Royet, J., Reichhart, J. M. & Levashina, E. A. Silencing of Toll pathway components by direct injection of double-stranded RNA into *Drosophila* adult flies. *Nucleic Acids Res.* **31**, 6619–6623 (2003).

# Regulatory T-cell suppressor program co-opts transcription factor IRF4 to control T<sub>H</sub>2 responses

Ye Zheng<sup>1,2†</sup>, Ashutosh Chaudhry<sup>1,2†</sup>, Arnold Kas<sup>1,2</sup>, Paul deRoos<sup>1,2†</sup>, Jeong M. Kim<sup>1,2</sup>, Tin-Tin Chu<sup>1,2</sup>, Lynn Corcoran<sup>4</sup>, Piper Treuting<sup>3</sup>, Ulf Klein<sup>5</sup> & Alexander Y. Rudensky<sup>1,2†</sup>

In the course of infection or autoimmunity, particular transcription factors orchestrate the differentiation of T<sub>H</sub>1, T<sub>H</sub>2 or T<sub>H</sub>17 effector cells, the responses of which are limited by a distinct lineage of suppressive regulatory T cells (T<sub>reg</sub>). T<sub>reg</sub> cell differentiation and function are guided by the transcription factor Foxp3, and their deficiency due to mutations in Foxp3 results in aggressive fatal autoimmune disease associated with sharply augmented T<sub>H</sub>1 and T<sub>H</sub>2 cytokine production<sup>1–3</sup>. Recent studies suggested that Foxp3 regulates the bulk of the Foxp3-dependent transcriptional program indirectly through a set of transcriptional regulators serving as direct Foxp3 targets<sup>4,5</sup>. Here we show that in mouse T<sub>reg</sub> cells, high amounts of interferon regulatory factor-4 (IRF4), a transcription factor essential for T<sub>H</sub>2 effector cell differentiation, is dependent on Foxp3 expression. We proposed that IRF4 expression endows T<sub>reg</sub> cells with the ability to suppress T<sub>H</sub>2 responses. Indeed, ablation of a conditional *Irf4* allele in T<sub>reg</sub> cells resulted in selective dysregulation of T<sub>H</sub>2 responses, IL4-dependent immunoglobulin isotype production, and tissue lesions with pronounced plasma cell infiltration, in contrast to the mononuclear-cell-dominated pathology typical of mice lacking T<sub>reg</sub> cells. Our results indicate that T<sub>reg</sub> cells use components of the transcriptional machinery, promoting a particular type of effector CD4<sup>+</sup> T cell differentiation, to efficiently restrain the corresponding type of the immune response.

T<sub>reg</sub> cell deficiency results in activation and expansion of CD4<sup>+</sup> and CD8<sup>+</sup> T cells, dendritic cells, granulocytes and macrophages, and greatly increased production of a wide range of cytokines including interleukin (IL)-2, T<sub>H</sub>1 and T<sub>H</sub>2 cytokines<sup>6,7</sup>. Expression of Foxp3 is required for the establishment and maintenance of T<sub>reg</sub> lineage identity and suppressor function<sup>8–11</sup>. Our recent study suggested that in T<sub>reg</sub> cells Foxp3 might regulate expression of IRF4 (refs 12–14) a transcription factor that is indispensable for T<sub>H</sub>2 effector cell differentiation<sup>15,16</sup>. Furthermore, a recent study suggested a prominent role for IRF4 in T<sub>H</sub>17 differentiation<sup>17</sup>. Thus, we decided to examine a role for IRF4 in T<sub>reg</sub> cell differentiation and function.

Foxp3 binding within the promoter region of *Irf4* in T<sub>reg</sub> cells<sup>4</sup> was confirmed by chromatin immunoprecipitation (ChIP)-coupled quantitative PCR (qPCR) (Supplementary Fig. 1a, b). *Irf4* messenger RNA was increased in thymic and peripheral Foxp3<sup>+</sup> T<sub>reg</sub> cells in comparison to CD25<sup>−</sup> Foxp3<sup>−</sup> CD4<sup>+</sup> T cells (data not shown)<sup>8</sup>. Furthermore, Foxp3 knockdown using a retrovirally encoded Foxp3-specific short hairpin RNA resulted in a marked diminution in *Irf4* mRNA (Supplementary Fig. 1c), suggesting that Foxp3 directly regulates IRF4 expression in T<sub>reg</sub> cells.

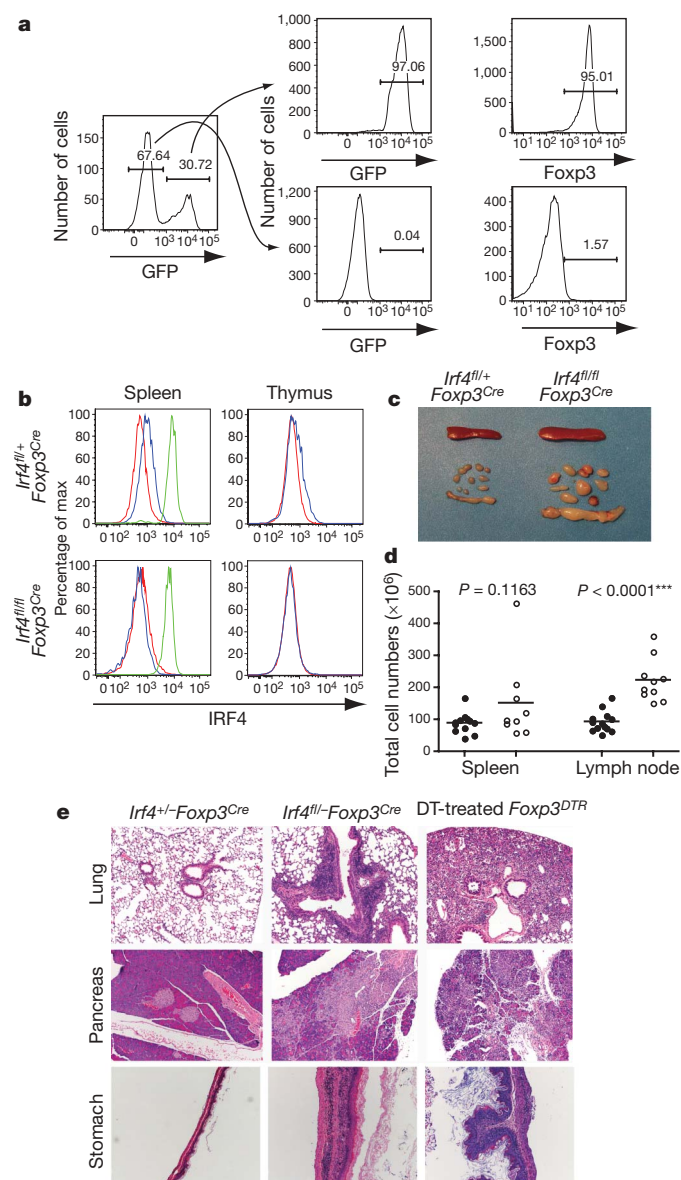
Next we induced deletion of a conditional *Irf4* allele (*Irf4*<sup>fl</sup>) in T<sub>reg</sub> cells by crossing *Irf4*<sup>fl</sup> mice to Foxp3<sup>Cre</sup> mice expressing yellow

fluorescent protein (YFP)-Cre recombinase fusion protein under the Foxp3 locus control<sup>18,19</sup>. The *Irf4*<sup>fl</sup> allele has a 'built-in' reporter capacity in that Cre-mediated recombination results in the deletion of the *Irf4* promoter and the exon containing the translational start, and the concomitant expression of green fluorescent protein (GFP)<sup>18</sup>. The specificity of the *Irf4*<sup>fl</sup> deletion was examined by flow cytometric analysis of Foxp3 expression in a sorted GFP<sup>+</sup> CD4<sup>+</sup> T cell population that underwent Cre-mediated recombination, and in a GFP<sup>−</sup> CD4<sup>+</sup> T cell population that did not. Essentially all GFP<sup>+</sup> cells expressed Foxp3, whereas GFP<sup>−</sup> cells lacked Foxp3 expression (Fig. 1a). Flow-cytometric analysis of control *Irf4*<sup>fl/+</sup> Foxp3<sup>Cre</sup> mice showed that IRF4 expression was markedly increased in all peripheral T<sub>reg</sub> cells, but only modestly in Foxp3<sup>+</sup> CD4<sup>+</sup> thymocytes, whereas IRF4 levels were undetectable in corresponding Foxp3<sup>−</sup> cell subsets. On Cre-mediated deletion in *Irf4*<sup>fl/fl</sup> Foxp3<sup>Cre</sup> mice, IRF4 protein became undetectable in both peripheral and thymic Foxp3<sup>+</sup> cells (Fig. 1b). *Irf4*<sup>fl/fl</sup> Foxp3<sup>Cre</sup> and *Irf4*<sup>fl/−</sup> Foxp3<sup>Cre</sup> mice harbouring an IRF4-deficient T<sub>reg</sub> subset were born at the expected Mendelian ratio and were indistinguishable from their wild-type or heterozygous littermates during the first month of life. However, by 6–8 weeks of age *Irf4*<sup>fl/fl</sup> Foxp3<sup>Cre</sup> and *Irf4*<sup>fl/−</sup> Foxp3<sup>Cre</sup> mice manifested identical autoimmune disease—including lymphadenopathy, weight loss, blepharitis and dermatitis—and succumbed to disease at 3–4 months of age (Fig. 1c, d and data not shown). Histopathological evaluation of diseased mice showed massive infiltration in the pancreas, lung and stomach, whereas control littermates did not show any noticeable pathology. For comparison to a complete T<sub>reg</sub> cell deficiency, we analysed tissue lesions in Foxp3<sup>DTR</sup> knock-in mice (expressing human diphtheria receptor, DTR under control of Foxp3 locus) that were subjected to chronic ablation of a T<sub>reg</sub> cell subset caused by diphtheria toxin treatment, starting from birth. These mice showed analogous lesions in the pancreas and stomach, and much more severe lesions in the lung in comparison to mice harbouring an IRF4-deficient T<sub>reg</sub> subset (Fig. 1e and data not shown). In contrast to the massive liver lesions observed after T<sub>reg</sub> depletion, livers in mice containing an IRF4-deficient T<sub>reg</sub> subset were unaffected, but kidneys showed the opposite trend (Supplementary Fig. 2 and data not shown). Furthermore, flow cytometric analysis showed expansion and activation of peripheral T cells but not dendritic cells in *Irf4*<sup>fl/fl</sup> Foxp3<sup>Cre</sup> mice, in contrast to the increased numbers of activated dendritic cells observed in T<sub>reg</sub>-deficient mice (Fig. 2a, b and data not shown).

The aforementioned phenotypic differences between mice lacking T<sub>reg</sub> cells and mice containing IRF4-deficient T<sub>reg</sub> cells could be due to a numerical decrease in the T<sub>reg</sub> cells in the absence of IRF4. However, the IRF4-deficient T<sub>reg</sub> subset was markedly increased in

<sup>1</sup>Howard Hughes Medical Institute, <sup>2</sup>Department of Immunology, <sup>3</sup>Department of Comparative Medicine, University of Washington, Seattle, Washington 98195, USA. <sup>4</sup>Immunology Division, The Walter and Eliza Hall Institute, Parkville Victoria 3050, Australia. <sup>5</sup>Institute for Cancer Genetics and Herbert Irving Comprehensive Cancer Center, Columbia University, New York, New York 10032, USA. †Present address: Department of Immunology, Memorial Sloan-Kettering Cancer Center, New York, New York 10021, USA.





**Figure 1 | Ablation of IRF4 in  $T_{reg}$  cells results in autoimmune lymphoproliferative disease.** **a**,  $Foxp3^{Cre}$ -mediated deletion of  $Irfl$  is restricted to  $T_{reg}$  cells. Spleen and lymph node  $CD4^{+}$  T cells from  $Irfl^{fl/fl} Foxp3^{Cre}$  mice were FACS-sorted into  $GFP^{+}$  and  $GFP^{-}$  cells containing a recombinant and unrecombined  $Irfl^{fl}$  allele, respectively, and stained for Foxp3. The post-sorting purity of  $GFP^{+}$  and  $GFP^{-}$  population was greater than 97%. **b**, Intracellular staining of IRF4 in splenic and thymic  $CD4^{+} Foxp3^{-}$  (red) and  $CD4^{+} Foxp3^{+}$  (blue) cells and splenic  $B220^{+} CD138^{+}$  (green) plasma cells (positive control) in  $Irfl^{fl/fl} Foxp3^{Cre}$  (top panels) and  $Irfl^{fl/fl} Foxp3^{Cre}$  mice (bottom panels). **c**, Splenomegaly and lymphadenopathy in  $Irfl^{fl/fl} Foxp3^{Cre}$  mice. **d**, Spleen and lymph node cellularity in  $Irfl^{fl/fl} Foxp3^{Cre}$  (filled circles) and  $Irfl^{fl/fl} Foxp3^{Cre}$  (open circles) mice. **e**, Histopathology induced after IRF4 ablation in  $T_{reg}$  cells. Representative haematoxylin and eosin (H&E)-stained tissue sections from 8-week-old  $Irfl^{fl/fl} Foxp3^{Cre}$  and  $Irfl^{fl/fl} Foxp3^{Cre}$  mice and diphtheria-toxin-treated  $Foxp3^{DTR}$  mice. Note pronounced infiltrates in the lung, pancreas and stomach of  $Irfl^{fl/fl} Foxp3^{Cre}$  mice and severe lesions in tissues from diphtheria-toxin-treated  $Foxp3^{DTR}$  mice ( $n = 3-5$ ). Original magnification (**c**, **e**),  $\times 10$ .

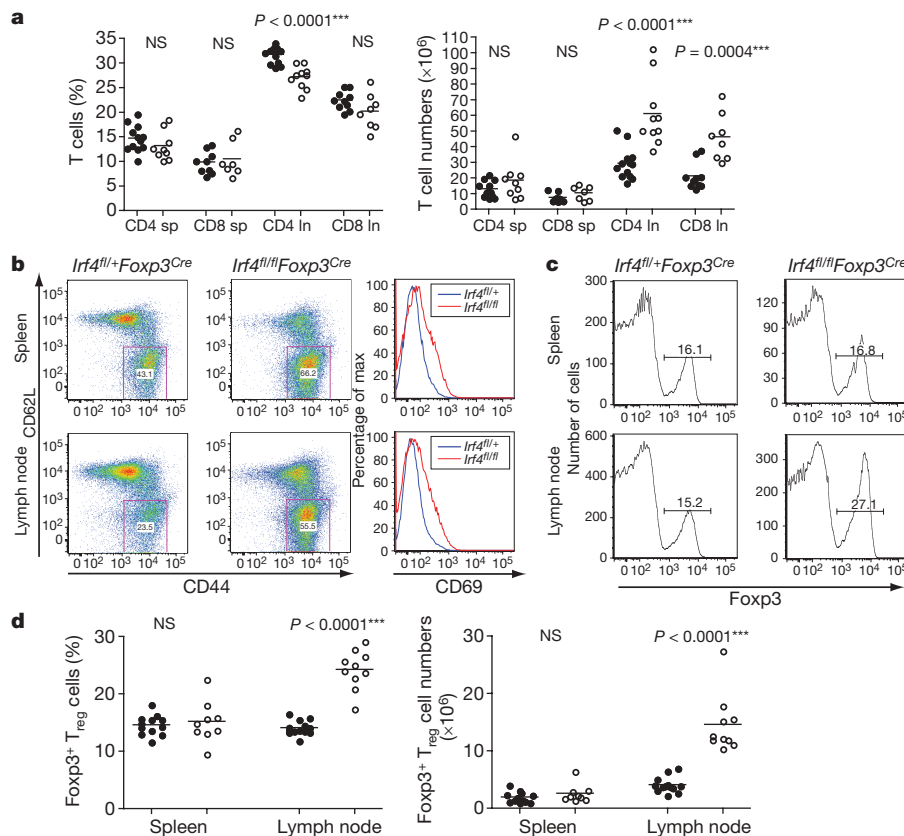
diseased  $Irfl^{fl/fl} Foxp3^{Cre}$  mice compared to  $Irfl^{fl/fl} Foxp3^{Cre}$  littermate controls, probably in response to T cell activation (Fig. 2c, d). IRF4 deficiency did not change Foxp3 protein expression on a per cell basis (Fig. 2c), excluding the possibility of down-modulation of Foxp3 in  $T_{reg}$  cells in the absence of IRF4 as a reason for their

functional incompetence. These results pointed to impairment in a particular aspect of  $T_{reg}$  function on IRF4 ablation.

Indeed, analysis of cytokine production by  $CD4^{+}$  T cells from  $Irfl^{fl/fl} Foxp3^{Cre}$  mice showed a notable increase in the numbers of IRF4-sufficient  $Foxp3^{-} CD4^{+}$  T cells producing the  $T_H2$  cytokines IL-4 and IL-5 (Fig. 3a, b), whereas IRF4-deficient  $Foxp3^{+} T_{reg}$  cells, unlike their IRF4-sufficient counterparts, were unable to produce effector cytokines (Supplementary Fig. 3). The production of IL-2 and the  $T_H1$  cytokine IFN- $\gamma$  by  $CD4^{+}$  T cells was largely unchanged, but production of the  $T_H17$  cytokine IL-17 was marginally increased in  $Irfl^{fl/fl} Foxp3^{Cre}$  mice compared to  $Irfl^{fl/+} Foxp3^{Cre}$  controls (Fig. 3a, b and data not shown). In agreement with these results, enzyme-linked immunosorbent assay (ELISA) showed high amounts of the  $T_H2$  cytokines IL-5, IL-13 and IL-10 in the supernatants of CD3-stimulated splenic  $Irfl^{fl/fl} Foxp3^{Cre}$  T cell cultures (Supplementary Fig. 4). In contrast, a lack of  $T_{reg}$  cells in  $Foxp3^{-}$  mice led to sharply increased production of IFN- $\gamma$  and IL-4 by  $CD4^{+}$  T cells (Supplementary Fig. 5). Massive increases in IFN- $\gamma$  and IL-2 production were also observed after ablation of  $T_{reg}$  cells in  $Foxp3^{DTR}$  mice<sup>7</sup> (J. M. Lund and A.Y.R., unpublished observations). Thus,  $T_H2$  effector  $CD4^{+}$  T cell responses seem to be selectively dysregulated in mice containing IRF4-deficient  $T_{reg}$  cells. Notably, the augmented  $T_H2$  responses appeared in  $Irfl^{fl/fl} Foxp3^{Cre}$  mice only by 3–4 weeks of age, whereas both  $T_H2$  and  $T_H1$  cytokine production in 11-day-old  $Irfl^{fl/fl} Foxp3^{Cre}$  and  $Irfl^{fl/+} Foxp3^{Cre}$  mice was low (data not shown).

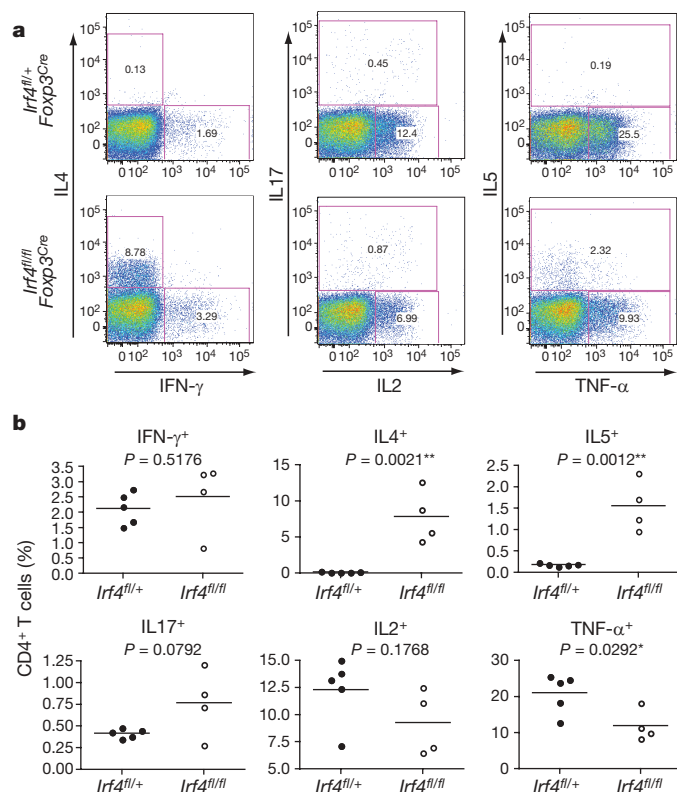
Because IL-4 is an important cytokine that promotes IgG1 and IgE class-switch recombination, we measured levels of serum immunoglobulin isotypes in diseased 8-week-old  $Irfl^{fl/fl} Foxp3^{Cre}$  and littermate control  $Irfl^{fl/+} Foxp3^{Cre}$  mice. As a positive control, we also analysed 3–4-week-old  $T_{reg}$ -deficient  $Foxp3^{-}$  mice. Consistent with an uncontrolled  $T_H2$  response, IL-4-dependent IgG1 and IgE serum concentrations reached very high levels ( $10-30 \text{ mg ml}^{-1}$  and  $0.75-1.5 \text{ mg ml}^{-1}$ , respectively) in  $Irfl^{fl/fl} Foxp3^{Cre}$  mice, whereas serum IgM and IgA concentrations were only modestly increased in comparison to control mice. However, the amounts of the remaining immunoglobulin isotypes, including IFN- $\gamma$ -dependent IgG2a, were diminished in the serum of afflicted mice (Fig. 4a). Unlike  $Irfl^{fl/fl} Foxp3^{Cre}$  mice,  $Foxp3^{-}$  mice showed in addition to increased IgG1 and IgE, sharply increased amounts of IgG2a consistent with the increased IFN- $\gamma$  production in these mice (Fig. 4a). Furthermore, we observed ubiquitous germinal centre formation and increased numbers of plasma cells in the spleens of diseased  $Irfl^{fl/fl} Foxp3^{Cre}$  mice, but not in the littermate controls (Fig. 4b and data not shown). In  $Irfl^{fl/fl} Foxp3^{Cre}$  mice, 15% and 67% of plasma cells produced IgE and IgG1, respectively, in contrast to 3% IgG1-producing and undetectable numbers of IgE-producing plasma cells in control mice (Supplementary Fig. 6). Furthermore, examination of diseased mice harbouring an IRF4-deficient  $T_{reg}$  cell subset revealed a predominance of cells with a characteristic plasma cell morphology infiltrating the pancreas, kidney and stomach (Fig. 4c and data not shown). In contrast, the pancreatic and other tissue infiltrates in  $Foxp3^{DTR}$  mice subjected to chronic  $T_{reg}$  ablation, or in  $Foxp3^{-}$  mice, consisted predominantly of macrophages, lymphocytes and neutrophils with very few plasma cells (Fig. 4c and data not shown). We also tested whether the dysregulated  $T_H2$  response in the presence of IRF4-deficient  $T_{reg}$  cells promoted autoantibody production. Indeed, serum IgG1 from affected  $Irfl^{fl/fl} Foxp3^{Cre}$  mice showed robust reactivity with several tissue antigens, whereas minimal reactivity was found in control animals (Supplementary Fig. 7).

The distinct immune activation profile and tissue lesions in diseased mice harbouring an IRF4-deficient  $T_{reg}$  cell subset were in agreement with the previously mentioned idea that only a part of the suppressor program was impaired in IRF4-deficient  $T_{reg}$  cells. In further support of this notion, the *in vitro* suppressor capacity of IRF4-deficient  $T_{reg}$  cells isolated from  $Irfl^{fl/fl} Foxp3^{Cre}$  mice was largely intact (Fig. 5a). Furthermore, cell surface amounts of two putative  $T_{reg}$  suppressor effector molecules CTLA4 and CD73 (also known



**Figure 2 | Increased numbers and activation of CD4<sup>+</sup> T cells in mice harbouring IRF4-deficient T<sub>reg</sub> cells.** **a**, CD4<sup>+</sup> and CD8<sup>+</sup> T cell numbers in the spleen (sp) and lymph nodes (ln) of *Ir4<sup>fl/fl</sup> Fopx3<sup>Cre</sup>* (open circles) mice and *Ir4<sup>fl/fl</sup> Fopx3<sup>Cre</sup>* littermate control mice. NS, not significant. **b**, Flow cytometric analysis of CD44, CD62L and CD69 expression on CD4<sup>+</sup> T cells in 8-week-old *Ir4<sup>fl/fl</sup> Fopx3<sup>Cre</sup>* mice and

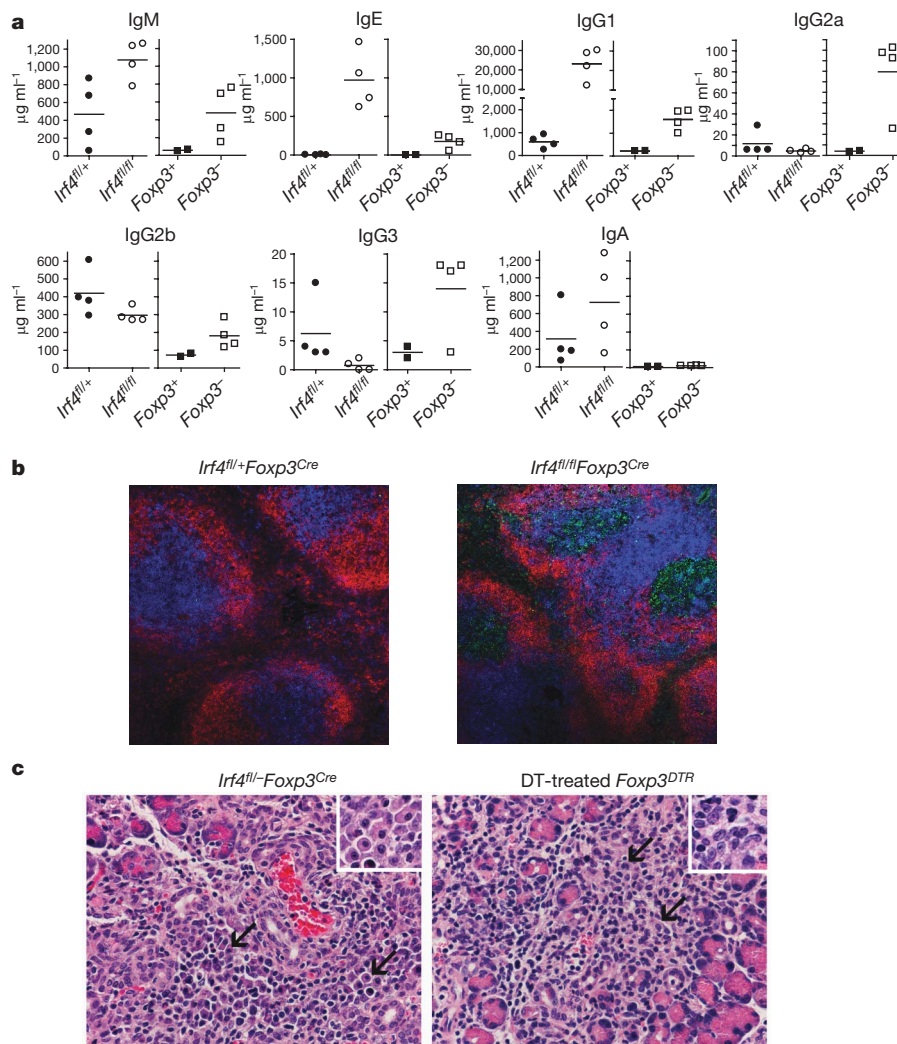
*Ir4<sup>fl/fl</sup> Fopx3<sup>Cre</sup>* littermates. **a**, A representative of three independent experiments is shown. **c**, **d**, Increased Fopx3<sup>+</sup> T<sub>reg</sub> cell subset in *Ir4<sup>fl/fl</sup> Fopx3<sup>Cre</sup>* mice. Flow cytometric analyses of spleen and lymph node cells from *Ir4<sup>fl/fl</sup> Fopx3<sup>Cre</sup>* mice (filled circles) and *Ir4<sup>fl/fl</sup> Fopx3<sup>Cre</sup>* mice (open circles) mice. A representative of three independent experiments is shown.



as NT5E) were not altered in IRF4-deficient T<sub>reg</sub> cells. The levels of CD25 (IL2RA) and GITR (TNFRSF18) were also unchanged (Fig. 5b and data not shown). However, ICOS expression was decreased, suggesting that IRF4 regulates expression of a distinct subset of functionally important genes in T<sub>reg</sub> cells.

An unbiased assessment of the effect of IRF4 deficiency on T<sub>reg</sub> transcriptome demonstrated that the expression of up to 20% of 'T<sub>reg</sub>-specific' genes decreased in the absence of IRF4, whereas approximately 7% were increased (Fig. 5c). Recent gene targeting studies demonstrated that T<sub>reg</sub> cells use several suppressor modalities<sup>19–25</sup>. Examination of the IRF4-dependent subset of 'T<sub>reg</sub>' genes and the subsequent independent confirmation by qPCR showed significantly decreased expression of several genes (*Fgl2*, *Il10* and *Gzmb*) encoding putative suppressor effector molecules; a few (*Ebi3* and *Entpd1*) were only marginally decreased, whereas others (*Tgfb1* and *Ctla4*) did not change (Fig. 5d, e and data not shown). It is likely that changes in expression of a combination of genes, but not a single gene, account for the impaired suppressor capacity of IRF4-deficient T<sub>reg</sub> cells. In this regard, we recently showed a role for IL10 in T<sub>reg</sub>-mediated suppression<sup>20</sup>. In addition, ICOS-deficient T<sub>reg</sub> cells were inferior in comparison to their wild-type counterparts in limiting

**Figure 3 | IRF4 deficiency in T<sub>reg</sub> cells results in a selective failure to control TH2 responses.** **a**, **b**, Flow cytometric analysis of cytokine production by splenic CD4<sup>+</sup> T cells from *Ir4<sup>fl/fl</sup> Fopx3<sup>Cre</sup>* mice and *Ir4<sup>fl/fl</sup> Fopx3<sup>Cre</sup>* littermates. Splenocytes were stimulated with CD3 (5 μg ml<sup>-1</sup>) and CD28 (5 μg ml<sup>-1</sup>) antibodies in the presence of Golgi-Plug (1 μg ml<sup>-1</sup>) for 5 h before staining for CD4, CD8 and the indicated cytokines. A representative of three independent experiments is shown.



**Figure 4 | Increased serum IgG1 and IgE concentration, germinal centre formation, and plasma cell tissue infiltration caused by IRF4 deficiency in T<sub>reg</sub> cells.** **a**, Analysis of immunoglobulin isotype amounts in sera of 8-week-old *Irf4<sup>fl/fl</sup>* *Foxp3<sup>Cre</sup>* mice and *Irf4<sup>fl/+</sup>* *Foxp3<sup>Cre</sup>* littermates, and of 3–4-week-old *Foxp3<sup>-</sup>* and *Foxp3<sup>+</sup>* littermates. **b**, Immunofluorescent staining of germinal centre B cells (GL7<sup>+</sup>, green), follicular B cells (IgD<sup>+</sup>, red), and CD4<sup>+</sup> T cells (blue) in spleens of *Irf4<sup>fl/+</sup>* *Foxp3<sup>Cre</sup>* mice and *Irf4<sup>fl/fl</sup>* *Foxp3<sup>Cre</sup>* littermates. Original magnification,  $\times 20$ . **c**, Histological sections of H&E-stained pancreas from 8-week-old *Irf4<sup>fl/-</sup>* *Foxp3<sup>Cre</sup>* mice and T<sub>reg</sub>-deficient

diphtheria-toxin-treated *Foxp3<sup>DTR</sup>* mice. The *Irf4<sup>fl/-</sup>* *Foxp3<sup>Cre</sup>* pancreas is infiltrated primarily by plasma cells (arrows), that is, distinct round cells containing an eccentric nucleus with a cartwheel chromatin appearance and perinuclear clearing (inset; original magnification,  $\times 60$ ). In contrast, the pancreatic infiltrates of diphtheria-toxin-treated *Foxp3<sup>DTR</sup>* mice contained principally macrophages (arrows), that is, large cells with abundant eosinophilic cytoplasm, reniform to oval nuclei, and indistinct cell borders (inset; original magnification,  $\times 60$ ). Original magnification for both panels,  $\times 40$ . Representative sections are shown.

expansion of effector T cells in lymphopenic hosts (Y.Z., unpublished observations).

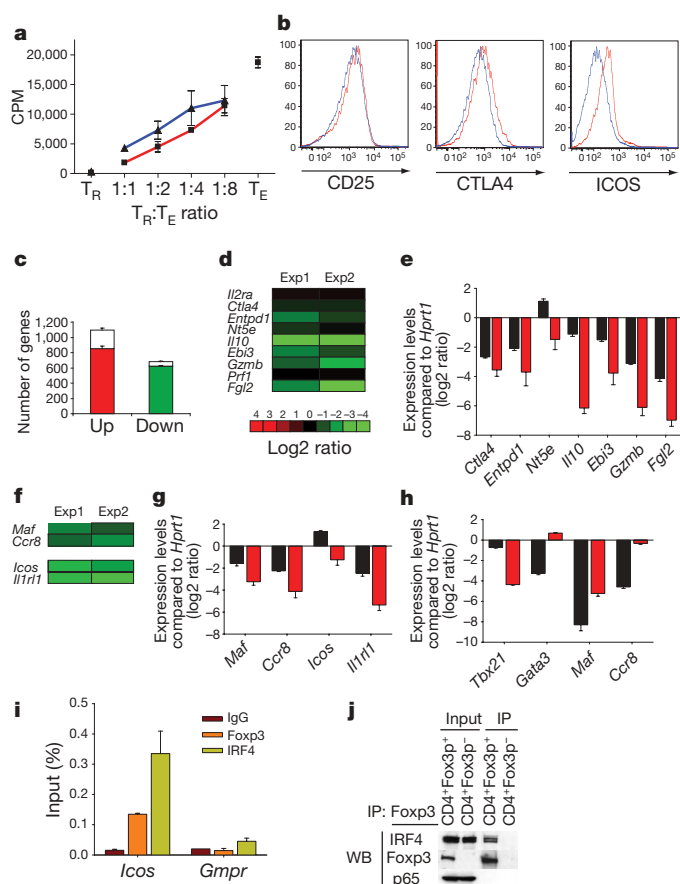
The observation that ICOS—the expression of which is critically important for T<sub>H2</sub> differentiation<sup>26</sup>—was expressed in T<sub>reg</sub> cells in an IRF4-dependent manner suggested that IRF4 might be involved in the regulation of a functionally important transcriptional module shared by both T<sub>reg</sub> and T<sub>H2</sub> effector T cells. To test this hypothesis, we cross-referenced the IRF4-dependent gene data set derived from our T<sub>reg</sub> cell studies to an effector T<sub>H2</sub>-specific gene data set. This comparison and independent qPCR analysis showed that in addition to *Icos*, the expression of two important T<sub>H2</sub> genes *Maf* and *Ccr8* was compromised in IRF4-deficient T<sub>reg</sub> cells, and the expression *Il1rl*—another essential player in T<sub>H2</sub> differentiation—was also decreased (Fig. 5f–h)<sup>27–29</sup>.

Next, using *Icos* as a model gene, we sought to investigate the possibility that in T<sub>reg</sub> cells IRF4 and Foxp3 co-operate in transcriptional regulation. Using CLOVER, a transcription factor binding site prediction algorithm<sup>30</sup>, we identified a putative IRF4 binding site

within the *Icos* promoter corresponding to the Foxp3 binding sites<sup>4</sup> (Supplementary Fig. 8), and confirmed binding of both Foxp3 and IRF4 by ChIP-coupled qPCR (Fig. 5i). These results raised the possibility that Foxp3 physically interacts with IRF4. In fact, this interaction was confirmed by IRF4 and Foxp3 co-immunoprecipitation from T<sub>reg</sub> cell nuclear lysates using a Foxp3-specific antibody (Fig. 5j). These results indicate that IRF4 binds to Foxp3 and the resulting complexes affect expression of certain target genes such as *Icos*.

Our results indicate that Foxp3 induces IRF4 expression in T<sub>reg</sub> cells. A transcriptional module downstream of IRF4 is probably modified after IRF4 interaction with Foxp3 to facilitate efficient T<sub>H2</sub> suppression by T<sub>reg</sub> cells. Uniformly increased IRF4 protein expression in peripheral T<sub>reg</sub> cells seems to suggest that they are equally poised to suppress T<sub>H2</sub> responses. However, T<sub>reg</sub> cell populations may still be heterogeneous in this regard, if signalling-dependent post-translational modifications of IRF4 or recruitment of other nuclear factors are needed for T<sub>reg</sub>-mediated T<sub>H2</sub> suppression. We propose that T<sub>reg</sub> cells might hijack certain components of





**Figure 5 | IRF4 interacts with Foxp3 and diminished expression of a subset of suppressor effector and  $T_H2$  specific genes in IRF4-deficient  $T_{reg}$  cells.**

**a**, IRF4-sufficient ( $Ir4^{+/+}$   $T_{reg}$ , red) and -deficient ( $Ir4^{-/-}$   $T_{reg}$ , blue)  $T_{reg}$  cells from 5-week-old  $Ir4^{+/+}$   $Foxp3^{Cre}$  mice and  $Ir4^{fl/fl}$   $Foxp3^{Cre}$  littermates suppress *in vitro* proliferative response of  $Foxp3^{-/-}$   $CD4^{+}$   $T$  cells ( $T_E$ ) from B6 mice. A representative of two independent experiments is shown. **b**, Flow cytometric analysis of CD25, CTLA4 and ICOS expression by IRF4-sufficient (red) and -deficient (blue)  $T_{reg}$  cells. **c**, Numbers of IRF4-independent genes that were up- (red bars) or downregulated (green bars), respectively, in  $T_{reg}$  cells compared to naive  $CD25^{-}$   $Foxp3^{-}$   $CD4^{+}$   $T$  cells<sup>9</sup>. Open bars represent genes in which expression was changed by twofold or more in the absence of IRF4. **d**, Decreased expression of genes with a presumed role in  $T_{reg}$  suppressor function in IRF4-deficient compared to IRF4-sufficient  $T_{reg}$  cells. The data (**c**, **d**) represent average of two independent microarray experiments (exp) performed using YFP- $Cre^{+}$   $T_{reg}$  cells FACS-purified from healthy  $Ir4^{fl/fl}$   $Foxp3^{Cre/wt}$  and  $Ir4^{+/+}$   $Foxp3^{Cre}$  littermates. **e**, qPCR analysis of relative expression of genes shown in **d** in  $Ir4^{+/+}$   $T_{reg}$  (black) and  $Ir4^{-/-}$   $T_{reg}$  (red). **f**, The decreased expression of  $T_H2$ -specific or functionally important genes in IRF4-deficient in comparison to IRF4-sufficient  $T_{reg}$  cells (two independent microarray experiments as above). **g**, **h**, qPCR analysis of relative expression of the  $T_H2$ -specific gene set in  $Ir4^{+/+}$   $T_{reg}$  (black) and  $Ir4^{-/-}$   $T_{reg}$  (red) (**g**), and *in vitro* differentiated  $T_H1$  (black) and  $T_H2$  cells (red) (**h**). Data in **e**, **g** and **h** represent mean and s.d. of the expression of genes relative to *Hprt1* in two independent experiments using three replicates each. **i**, Both IRF4 and Foxp3 bind to the promoter region of the *Icos* gene. qPCR analysis of Foxp3- and IRF4-bound chromatin isolated from wild-type  $T_{reg}$  cells using primer set corresponding to the *Icos* promoter region. IgG ChIP and qPCR using primers corresponding to the promoter region of *Gmpr* was used as a specificity controls. **j**, Western blot (WB) analysis of IRF4 in nuclear lysates of wild-type  $Foxp3^{+}$   $T_{reg}$  cells and total  $Foxp3^{-}$   $CD4^{+}$   $T$  cells (control) (lanes 1 and 2), and in Foxp3 complexes immunoprecipitated (IP) from the nuclear lysates using Foxp3 antibody. Transcription factor p65 was a negative control. IRF4 signal in control nuclear lysates is due to the presence of activated IRF4 $^{+}$   $CD25^{+}$   $Foxp3^{-}$   $T$  cells.

transcriptional machinery promoting a particular effector T cell differentiation to efficiently control the corresponding type of the immune response.

## METHODS SUMMARY

**Mice.** Mice containing  $Ir4^{fl}$ ,  $Ir4^{-/-}$ ,  $Foxp3^{Cre}$  and  $Foxp3^{DTR}$  alleles were previously described<sup>17,18,19</sup>. The  $Ir4^{fl/fl}$   $Foxp3^{Cre}$  and  $Ir4^{fl/-}$   $Foxp3^{Cre}$  mice were indistinguishable in regards to the IRF4 deletion efficiency in  $T_{reg}$  cells and to the immunological and clinical manifestations of autoimmunity, and so were all three types of healthy littermate control mice,  $Ir4^{+/+}$   $Foxp3^{Cre}$ ,  $Ir4^{fl/+}$   $Foxp3^{Cre}$  and  $Ir4^{+/+}$   $Foxp3^{Cre}$ .

**Flow cytometric and serum immunoglobulin analyses.** Rat anti-mouse IRF4 monoclonal antibody (clone 3E4) was raised against a glutathione S-transferase (GST)-fusion protein containing the carboxy-terminal 65-amino-acid sequence of murine IRF4. FACS data were acquired on a FACSCanto flow cytometer (Becton Dickinson) and analysed using FlowJo software package (Tri-Star).

Serum IgM, IgG1, IgG2a, IgG2b, IgG3 and IgA concentrations were measured using SBA Clonotyping System (Southern Biotech). IgE ELISA was performed using biotinylated anti-IgE antibody (BD Pharmingen) and streptavidin-conjugated HRP.

**ChIP and qPCR.** ChIP with Foxp3 and IRF4 antibodies (Santa Cruz Biotechnology) were performed using MACS-purified  $CD4^{+}$   $CD25^{+}$   $T_{reg}$  cells as previously described<sup>4</sup>. qPCRs were performed using primers listed in Supplementary Table 1.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 18 February; accepted 21 November 2008.

Published online 1 February 2009.

- Ramsdell, F. Foxp3 and natural regulatory T cells: key to a cell lineage? *Immunity* **19**, 165–168 (2003).
- Sakaguchi, S. et al. Foxp3 $^{+}$   $CD25^{+}$   $CD4^{+}$  natural regulatory T cells in dominant self-tolerance and autoimmune disease. *Immunol. Rev.* **212**, 8–27 (2006).
- Zheng, Y. & Rudensky, A. Y. Foxp3 in control of the regulatory T cell lineage. *Nature Immunol.* **8**, 457–462 (2007).
- Zheng, Y. et al. Genome-wide analysis of Foxp3 target genes in developing and mature regulatory T cells. *Nature* **445**, 936–940 (2007).
- Marson, A. et al. Foxp3 occupancy and regulation of key target genes during T-cell stimulation. *Nature* **445**, 931–935 (2007).
- Kanangat, S. et al. Disease in the scurfy (sf) mouse is associated with overexpression of cytokine genes. *Eur. J. Immunol.* **26**, 161–165 (1996).
- Kim, J. M., Rasmussen, J. P. & Rudensky, A. Y. Regulatory T cells prevent catastrophic autoimmunity throughout the lifespan of mice. *Nature Immunol.* **8**, 191–197 (2007).
- Fontenot, J. D. et al. Regulatory T cell lineage specification by the forkhead transcription factor foxp3. *Immunity* **22**, 329–341 (2005).
- Gavin, M. A. et al. Foxp3-dependent programme of regulatory T-cell differentiation. *Nature* **445**, 771–775 (2007).
- Lin, W. et al. Regulatory T cell development in the absence of functional Foxp3. *Nature Immunol.* **8**, 359–368 (2007).
- Williams, L. M. & Rudensky, A. Y. Maintenance of the Foxp3-dependent developmental program in mature regulatory T cells requires continued expression of Foxp3. *Nature Immunol.* **8**, 277–284 (2007).
- Eisenbeis, C. F., Singh, H. & Storb, U. Pip, a novel IRF family member, is a lymphoid-specific, PU.1-dependent transcriptional activator. *Genes Dev.* **9**, 1377–1387 (1995).
- Mittrucker, H. W. et al. Requirement for the transcription factor LSIRF/IRF4 for mature B and T lymphocyte function. *Science* **275**, 540–543 (1997).
- Iida, S. et al. Deregulation of MUM1/IRF4 by chromosomal translocation in multiple myeloma. *Nature Genet.* **17**, 226–230 (1997).
- Lohoff, M. et al. Dysregulated T helper cell differentiation in the absence of interferon regulatory factor 4. *Proc. Natl Acad. Sci. USA* **99**, 11808–11812 (2002).
- Rengarajan, J. et al. Interferon regulatory factor 4 (IRF4) interacts with NFATc2 to modulate interleukin 4 gene expression. *J. Exp. Med.* **195**, 1003–1012 (2002).
- Brustle, A. et al. The development of inflammatory  $T_H17$  cells requires interferon-regulatory factor 4. *Nature Immunol.* **8**, 958–966 (2007).
- Klein, U. et al. Transcription factor IRF4 controls plasma cell differentiation and class-switch recombination. *Nature Immunol.* **7**, 773–782 (2006).
- Rubtsov, Y. P. et al. Regulatory T cell-derived interleukin-10 limits inflammation at environmental interfaces. *Immunity* **28**, 546–558 (2008).
- Collison, L. W. et al. The inhibitory cytokine IL-35 contributes to regulatory T-cell function. *Nature* **450**, 566–569 (2007).
- Li, M. O., Wan, Y. Y. & Flavell, R. A. T cell-produced transforming growth factor- $\beta$  controls T cell tolerance and regulates Th1- and Th17-cell differentiation. *Immunity* **26**, 579–591 (2007).
- Takahashi, T. et al. Immunologic self-tolerance maintained by  $CD25^{+}$   $CD4^{+}$  regulatory T cells constitutively expressing cytotoxic T lymphocyte-associated antigen 4. *J. Exp. Med.* **192**, 303–310 (2000).

23. Kobie, J. J. *et al.* T regulatory and primed uncommitted CD4 T cells express CD73, which suppresses effector CD4 T cells by converting 5'-adenosine monophosphate to adenosine. *J. Immunol.* **177**, 6780–6786 (2006).
24. Deaglio, S. *et al.* Adenosine generation catalyzed by CD39 and CD73 expressed on regulatory T cells mediates immune suppression. *J. Exp. Med.* **204**, 1257–1265 (2007).
25. Cao, X. *et al.* Granzyme B and perforin are important for regulatory T cell-mediated suppression of tumor clearance. *Immunity* **27**, 635–646 (2007).
26. Nurieva, R. I. *et al.* Transcriptional regulation of Th2 differentiation by inducible costimulator. *Immunity* **18**, 801–811 (2003).
27. Ho, I. C., Lo, D. & Glimcher, L. H. c-maf promotes T helper cell type 2 (Th2) and attenuates Th1 differentiation by both interleukin 4-dependent and -independent mechanisms. *J. Exp. Med.* **188**, 1859–1866 (1998).
28. Chensue, S. W. *et al.* Aberrant *in vivo* T helper type 2 cell response and impaired eosinophil recruitment in CC chemokine receptor 8 knockout mice. *J. Exp. Med.* **193**, 573–584 (2001).
29. Lohning, M. *et al.* T1/ST2 is preferentially expressed on murine Th2 cells, independent of interleukin 4, interleukin 5, and interleukin 10, and important for Th2 effector function. *Proc. Natl Acad. Sci. USA* **95**, 6930–6935 (1998).
30. Friith, M. C. *et al.* Detection of functional DNA motifs via statistical over-representation. *Nucleic Acids Res.* **32**, 1372–1381 (2004).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank K. Hilgner and K. Murphy for providing T<sub>H</sub>1, T<sub>H</sub>2 and T<sub>H</sub>0 gene expression data sets, B. Sullivan, R. Locksley, S. Quezada and J. Allison for critical reagents, K. Forbush and L. Karpik for expert technical assistance and mouse colony management, and R. Dalla-Favera for discussions. This work was supported by grants from the National Institutes of Health (A.Y.R.). Y.Z. and J.M.K. were supported by the CRI-Irvington Institute postdoctoral fellowship. A.Y.R. is an investigator with the Howard Hughes Medical Institute.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to A.Y.R. ([rudenska@mskcc.org](mailto:rudenska@mskcc.org)).

## METHODS

**Mice.** C57BL/6 (B6) mice were purchased from the Jackson Laboratory. Mice containing *Irf4<sup>fl</sup>*, *Irf4<sup>-/-</sup>*, *Foxp3<sup>Cre</sup>* and *Foxp3<sup>DTTR</sup>* alleles were previously described<sup>7,18,19</sup>. The *Irf4<sup>fl/fl</sup>* *Foxp3<sup>Cre</sup>* mice were indistinguishable from *Irf4<sup>fl/-</sup>* *Foxp3<sup>Cre</sup>* mice in regards to the efficiency of IRF4 deletion in T<sub>reg</sub> cells, and to the immunological and clinical manifestations of autoimmunity, as were all three types of healthy littermate control mice, *Irf4<sup>+/+</sup>* *Foxp3<sup>Cre</sup>*, *Irf4<sup>fl/+</sup>* *Foxp3<sup>Cre</sup>* and *Irf4<sup>+/+</sup>* *Foxp3<sup>Cre</sup>*. Mice were housed under specific pathogen-free conditions and used according to the guidelines of the Institutional Animal Care Committee at the University of Washington.

**Antibodies and FACS analysis.** Fluorescent-dye-conjugated antibodies were purchased from BD Pharmingen and eBioscience. Rat anti-mouse IRF4 monoclonal antibody (clone 3E4) was raised against a GST-fusion protein containing the C-terminal 65-amino-acid sequence of murine IRF4. FACS data were acquired on a FACSCanto flow cytometer (Becton Dickinson) and analysed using FlowJo software package (Tri-Star). Intracellular staining of Foxp3 and IRF4 was conducted using Foxp3 Mouse Regulatory T cell Staining Kit (eBioscience). For flow cytometric analysis of cytokine- and immunoglobulin-secreting cells, cell populations were first stained with antibodies against the indicated cell surface markers, followed by permeabilization in Fix/Perm buffer, and intracellular staining in Perm/Wash buffer (BD Pharmingen).

**Immunofluorescence microscopy.** Frozen tissue sections were fixed in cold acetone for 20 min, washed twice and pre-blocked in PBS containing 5% normal rat serum, 5% normal rabbit serum and 5% BSA for 30 min. Slides were then incubated in primary antibodies (anti-GL7-FITC, anti-IgD-biotin, anti-CD4-AF647) for 45 min, washed twice, and incubated with secondary antibodies (anti-FITC-AF488, StrepAv-AF555) for 45 min and fluorescence was examined using a Leica SL confocal microscope.

**In vitro suppression assay.** For *in vitro* suppression assays, T<sub>reg</sub> cells and 'effector' T cells were purified by positive selection using CD4-specific MACS beads (Miltenyi), followed by sorting on a FACSARIA cell sorter (Becton Dickinson). Antigen-presenting cells were prepared from wild-type B6 splenocytes by T-cell depletion using Thy1-specific MACS beads. Effector T cells ( $2 \times 10^4$  cells well<sup>-1</sup>) were co-cultured with T<sub>reg</sub> cells at indicated ratios in the presence of irradiated (30 Gy) antigen-presenting cells ( $1 \times 10^5$  cells well<sup>-1</sup>) in 96-well plates in complete RPMI1640 medium supplemented with 10% FBS and CD3 antibody (1 µg ml<sup>-1</sup>) for 60 h. One microcurie <sup>3</sup>H-thymidine was added

into the cultures for a further 8–12 h and cell proliferation in triplicate cultures was measured using a scintillation counter.

**Serum immunoglobulin ELISA.** Serum IgM, IgG1, IgG2a, IgG2b, IgG3 and IgA concentrations were measured using SBA Clonotyping System (Southern Biotech). For IgE, ELISA was performed using biotinylated anti-IgE detection antibody (BD Pharmingen) and streptavidin-conjugated HRP.

**Affymetrix microarray and qRT-PCR.** Total RNA was extracted with RNA Stat-60 reagent (Iso-Tex Diagnostics) from IRF4-sufficient and -deficient CD4<sup>+</sup> GFP/YFP-Cre<sup>+</sup> T<sub>reg</sub> cells FACS-purified from healthy *Irf4<sup>+/+</sup>* *Foxp3<sup>Cre</sup>* and *Irf4<sup>fl/-</sup>* *Foxp3<sup>Cre/wt</sup>* mice, respectively. Complementary DNA was synthesized using Superscript II Reverse Transcriptase (Invitrogen), and amplified twice and labelled using MessageAmp II aRNA Amplification kit (Ambion). Biotinylated complementary RNA was fragmented and hybridized to Affymetrix GeneChip Mouse Genome 430 2.0 arrays at the Stanford PAN Facility. All data analyses were performed by using Bioconductor for the statistical software R (<http://www.r-project.org>). Expression values were background corrected, normalized and summarized by using the default settings of the gcrma package. From the resulting data sets we extracted a list of genes with a more than twofold change in expression and cross-referenced this list to the previously published set of T<sub>reg</sub>-specific genes<sup>9</sup>. Effector T<sub>H2</sub>-specific gene data set was extracted by subtracting gene expression values for CD4 T cells activated *in vitro* under T<sub>H1</sub> skewing conditions or without skewing (T<sub>H0</sub>) from those for T cells activated under T<sub>H2</sub> skewing conditions (K. Hilgner and K. Murphy, unpublished observations).

To verify expression array data, independently prepared T<sub>reg</sub> cell samples were used to generate cDNA. qPCR was performed using Power SYBR Green PCR master mix (Applied Biosystems). PCR primer sequences are listed in Supplementary Table 1.

**Chromatin immunoprecipitation and qPCR.** Foxp3 and IRF4 antibody (Santa Cruz Biotechnology) ChIPs were performed using MACS purified CD4<sup>+</sup> CD25<sup>+</sup> T<sub>reg</sub> cells as previously described<sup>4</sup>. qPCRs were performed using primers listed in Supplementary Table 1.

**Immunoprecipitation and western blot analysis.** Nuclear extracts from sorted CD4<sup>+</sup> Foxp3<sup>+</sup> and CD4<sup>+</sup> Foxp3<sup>-</sup> cells were prepared in nuclear lysis buffer (Active Motif) according to the manufacturer's protocol. Immunoprecipitation was carried out using anti-Foxp3-coated magnetic beads followed by western blotting with Foxp3, IRF4 or p65 antibodies.



# Defensin-like polypeptide LUREs are pollen tube attractants secreted from synergid cells

Satohiro Okuda<sup>1\*</sup>, Hiroki Tsutsui<sup>1\*</sup>, Keiko Shiina<sup>1</sup>, Stefanie Sprunck<sup>2</sup>, Hidenori Takeuchi<sup>1</sup>, Ryoko Yui<sup>1</sup>, Ryushiro D. Kasahara<sup>1</sup>, Yuki Hamamura<sup>1</sup>, Akane Mizukami<sup>1</sup>, Daichi Susaki<sup>1</sup>, Nao Kawano<sup>1</sup>, Takashi Sakakibara<sup>1</sup>, Shoko Namiki<sup>1</sup>, Kie Itoh<sup>3</sup>, Kurataka Otsuka<sup>4</sup>, Motomichi Matsuzaki<sup>4</sup>, Hisayoshi Nozaki<sup>4</sup>, Tsuneyoshi Kuroiwa<sup>5</sup>, Akihiko Nakano<sup>4,6</sup>, Masahiro M. Kanaoka<sup>1</sup>, Thomas Dresselhaus<sup>2</sup>, Narie Sasaki<sup>1\*</sup> & Tetsuya Higashiyama<sup>1\*</sup>

For more than 140 years, pollen tube guidance in flowering plants has been thought to be mediated by chemoattractants derived from target ovules<sup>1</sup>. However, there has been no convincing evidence of any particular molecule being the true attractant that actually controls the navigation of pollen tubes towards ovules. Emerging data indicate that two synergid cells on the side of the egg cell emit a diffusible, species-specific signal to attract the pollen tube at the last step of pollen tube guidance<sup>1–3</sup>. Here we report that secreted, cysteine-rich polypeptides (CRPs) in a subgroup of defensin-like proteins are attractants derived from the synergid cells. We isolated synergid cells of *Torenia fournieri*, a unique plant with a protruding embryo sac, to identify transcripts encoding secreted proteins as candidate molecules for the chemoattractant(s). We found two CRPs, abundantly and predominantly expressed in the synergid cell, which are secreted to the surface of the egg apparatus. Moreover, they showed activity *in vitro* to attract competent pollen tubes of their own species and were named as LUREs. Injection of morpholino antisense oligomers against the LUREs impaired pollen tube attraction, supporting the finding that LUREs are the attractants derived from the synergid cells of *T. fournieri*.

Because the attractants derived from the synergid cells show rapid molecular evolution<sup>4</sup>, we decided to investigate peptides/proteins expressed in the synergid cells. We first developed a protocol to isolate synergid cells manually. The protruding embryo sacs of *T. fournieri* are especially suitable to isolate synergid cells as protoplasts (Fig. 1a)<sup>5</sup>. The synergid cell can be readily distinguished from other cells of the embryo sac (Fig. 1a and Supplementary Fig. 1a). After poly(A)<sup>+</sup> RNA extraction, reverse transcription and a 30-cycle long-distance polymerase chain reaction (PCR), we established a complementary DNA library from 25 synergid cells.

Expressed sequence tag (EST) and cluster analysis from 2,112 randomly picked clones generated 256 contigs (Supplementary Fig. 1b and Supplementary Table 1). Interestingly, we found that transcripts encoding CRPs—cysteine-rich small proteins with a putative signal peptide<sup>6</sup>—were most abundant in the cDNA library. Among the eight contigs with the highest number of ESTs, seven represented genes for CRPs. In total, we identified 16 CRPs with a putative signal peptide at the amino-terminal region (Supplementary Table 2; 6.3% of 256 contigs). These 16 CRPs represented ~29% of the total number of EST sequences. The 16 CRPs were numbered as TfCRP1–16 (*T. fournieri* cysteine-rich polypeptide 1–16) according to the size of the contigs. Because genes for CRPs are known to

diverge from each other rapidly<sup>6</sup>, it was difficult to find orthologues in other plants by molecular phylogenetic analysis.

The expression of many kinds of CRPs from synergid cells has also been reported as a characteristic feature in *Arabidopsis*<sup>7,8</sup> and maize<sup>9,10</sup>. Genome-wide analysis in *Arabidopsis* ovules has indicated that 56 CRPs were downregulated in a T-DNA insertion mutant of *MYB98* (ref. 8), which encodes an R2R3-MYB transcription factor specifically expressed in the synergid cell<sup>7,11</sup>. *MYB98* is necessary for both normal development of finger-like projections of the synergid cell wall (filiform apparatus) and pollen tube guidance<sup>11</sup>. Some of the CRPs predominantly expressed in the synergid cells have been shown to be secreted towards the filiform apparatus using green fluorescent protein (GFP) fusion<sup>7</sup>.

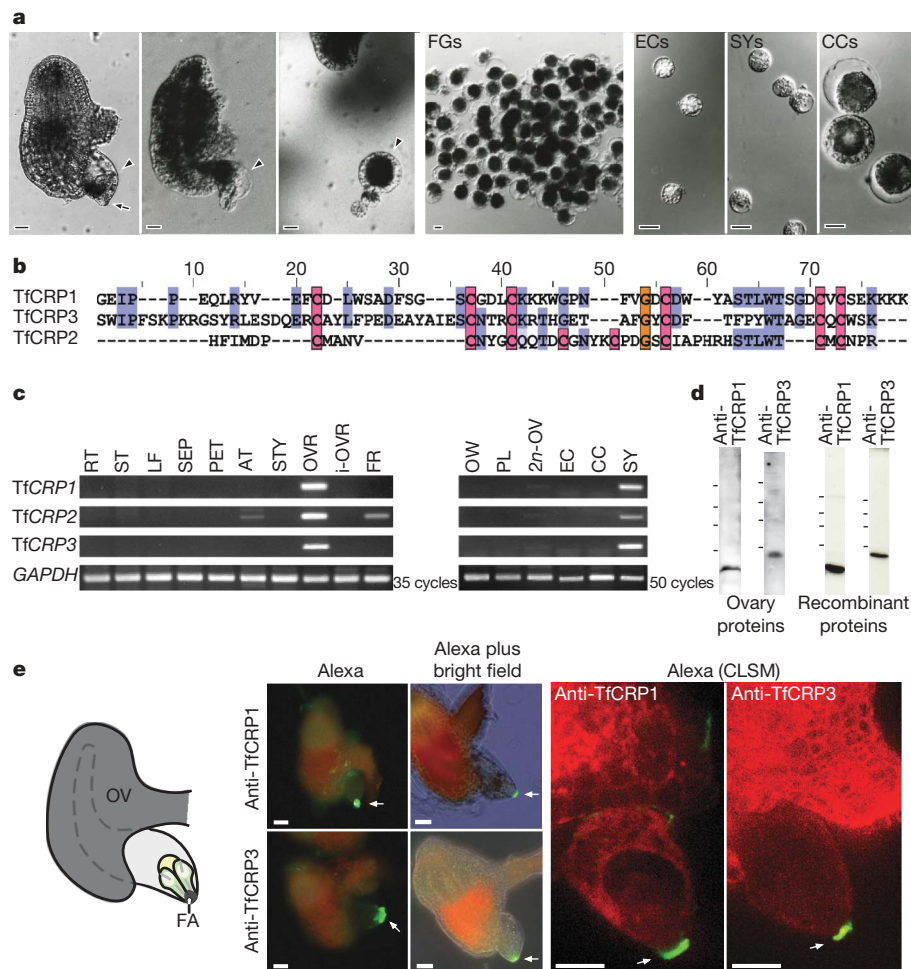
We focused our investigation on the three largest CRP contigs (TfCRP1–3; Fig. 1b), which contained more than 200 ESTs (Supplementary Fig. 1c). Abundant expression of these TfCRPs was confirmed using quantitative reverse transcription–PCR (qRT–PCR; Supplementary Fig. 1d). They were related to defensin-like proteins (DEFLs)<sup>12</sup>, as they contained CS $\alpha$  $\beta$ <sup>13</sup> and  $\gamma$ -core<sup>14</sup> motifs common to defensins and all classes of cysteine-containing antimicrobial peptides, respectively (Supplementary Table 2). Although CRPs are highly polymorphic<sup>6</sup>, putative mature TfCRP1 and TfCRP3, both of which contain 6 cysteines, are closely related according to the alignment of cysteine residues (Fig. 1b), their overall sequence similarity (62.1%), their acidic pK<sub>a</sub> (4.7 and 5.7) and the total number of amino acids (62 and 70; Supplementary Table 2). In contrast, TfCRP2 contains 8 cysteines and consists of a smaller number of amino acids (48).

The three CRPs of *T. fournieri* were specifically expressed in floral organs, as investigated by RT–PCR (Fig. 1c). TfCRP1 and TfCRP3 were expressed specifically in the mature ovary, predominantly in the synergid cell. TfCRP2 was additionally expressed in the anther and developing fruit, but predominantly in the synergid cell within the mature ovary.

We performed immunological analyses using antibodies against TfCRP1 and TfCRP3 (Fig. 1d, e). Consistent with the RT–PCR data analysis, TfCRP1 and TfCRP3 peptides were detected in the mature ovary of unpollinated pistils by immunoblotting analysis (Fig. 1d). Apparent molecular masses were 8.6 kD and 9.8 kD, respectively, which were similar to those of recombinant proteins—predicted mature polypeptide lacking N-terminal signal sequences (Fig. 1d; SignalP 3.0; Supplementary Table 2). This implied that the polypeptides were cleaved as predicted and were secreted. Consistently, these

<sup>1</sup>Division of Biological Science, Graduate School of Science, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Aichi 464-8602, Japan. <sup>2</sup>Cell Biology/Plant Physiology, University of Regensburg, Universitätsstraße 31, 93053 Regensburg, Germany. <sup>3</sup>Department of Integrated Biosciences, Graduate School of Frontier Sciences, University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8562, Japan. <sup>4</sup>Department of Biological Sciences, Graduate School of Science, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. <sup>5</sup>Research Information Center of Extremophile, Graduate School of Science, Rikkyo (Saint Paul's) University, Nishi-Ikebukuro, Tokyo 171-8501, Japan. <sup>6</sup>Molecular Membrane Biology Laboratory, RIKEN Advanced Science Institute, Hirosawa, Wako, Saitama 351-0198, Japan.

\*These authors contributed equally to this work.



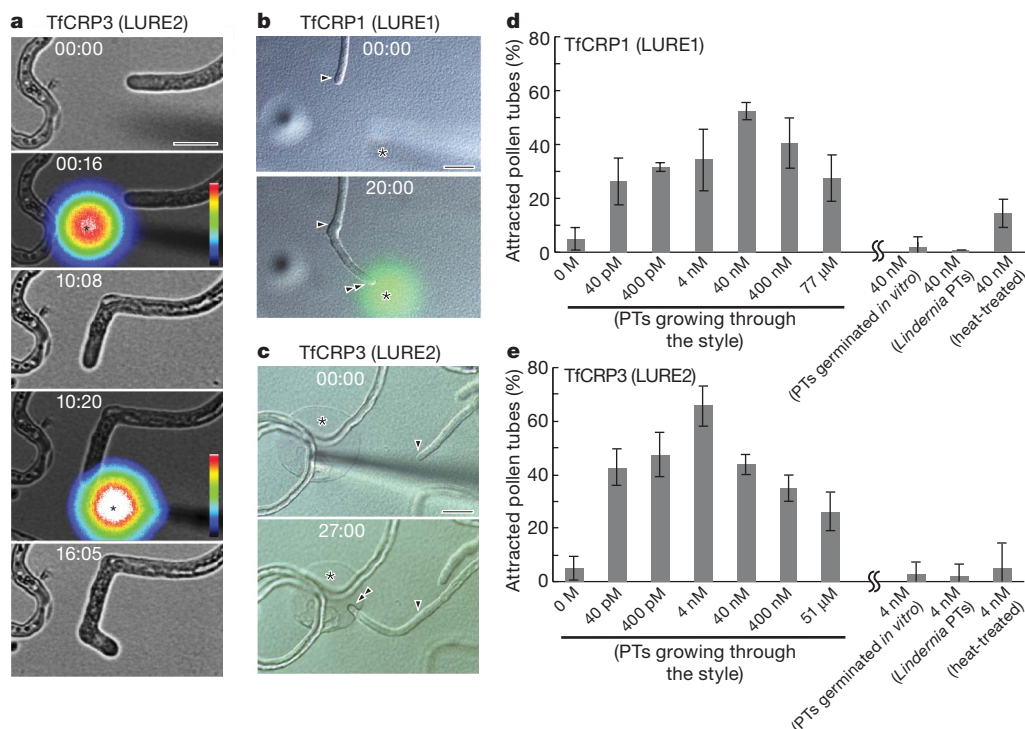
**Figure 1 | Identification and expression analysis of major CRPs of the synergid cell.** **a**, Ovules of *T. fournieri* were treated with a cell-wall-degrading enzyme solution to isolate the protruding embryo sac cells (arrowheads) as protoplasts (left three panels). An arrow indicates the micropylar end of the embryo sac to which a pollen tube is attracted. After collecting protoplasts of the female gametophytic cells (FGs), protoplasts of egg cells (ECs), synergid cells (SYs) and central cells (CCs) were further separated manually. Scale bars, 20  $\mu$ m. **b**, Alignment of three major CRPs identified by the EST-cluster analysis of the synergid cell. Only predicted mature CRPs are shown. Cysteine residues are labelled in pink, the conserved glycine residue of the defensin  $\gamma$ -core motif in orange, and other residues conserved in at least two of the three sequences in blue. **c**, RT-PCR analysis of organs/tissues of whole plants (left) and tissues/cells of ovaries (right). RT, root; ST, stem; LF, leaf; SEP, sepal; PET, petal; AT, anther; STY, style with a stigma; OVR, ovary; i-OVR, immature ovary with non-cellularized embryo sacs; FR, developing fruit containing heart-stage embryos; OW, ovary wall; PL, placenta; 2n-OV, sporophytic cells of ovaries;

GAPDH, glyceraldehyde-3-phosphate dehydrogenase. **d**, Immunoblotting analysis of *T. fournieri* ovary and recombinant proteins using the specific antibodies indicated. Recombinant TfCRP1 and TfCRP3 of predicted mature polypeptides lacking N-terminal signal sequences were used. The apparent molecular weights of TfCRP1 and TfCRP3 were 8.6 kD and 9.8 kD, respectively, in both analysis. Bars indicate position of molecular weight markers: 10.9, 16.2, 19.3 and 28.0 kD from the bottom. **e**, Schematic of the *T. fournieri* ovule (OV) containing a protruding embryo sac, and immunostaining showing CRP peptides secreted to the surface of the micropylar end of the embryo sac (green Alexa Fluor fluorescence, arrows). Alexa plus bright field images of ovules better illustrate the location of CRPs relative to the ovule structure. As indicated by the schematic model, the micropylar end of the embryo sac is occupied by a filiform apparatus (FA) formed by two synergid cells (green). The egg cell (yellow) is partly hidden behind synergid cells. CLSM, confocal laser scanning microscopy. Scale bars, 20  $\mu$ m.

two CRPs were detected at the micropylar surface of the filiform apparatus of the synergid cell by immunostaining of the ovule (Fig. 1e; Supplementary Fig. 1e). Properties of TfCRP1 and TfCRP3—namely, abundant and predominant expression in the synergid cell and small and secreted proteins, as well as rapid molecular evolution generally observed in CRPs<sup>6</sup>—all suggested that they might be pollen tube attractants derived from the synergid cell.

Next we investigated whether recombinant CRPs have the ability to attract pollen tubes (Fig. 2). Carboxy-terminal regions of TfCRP1 and TfCRP3 following the predicted cleavage site were expressed in *Escherichia coli*. Because TfCRP2 could not be expressed in *E. coli*, it was chemically synthesized. Polypeptides were refolded, implying that disulphide bonds were appropriately formed between cysteine residues. First, a pollen tube attraction assay was established using micropipettes (Fig. 2a and Supplementary Movie 1). When TfCRP3

was injected in front of a pollen tube, we found that 60% ( $n = 20$ ) of the tubes turned sharply towards the pipette; this never occurred when buffer alone was injected (0%,  $n = 20$ ). To quantitatively investigate the activity of the three peptides for pollen tube attraction, we developed a bead method, whereby peptides were embedded in gelatin beads and placed in front of the pollen tube ( $\sim 50$   $\mu$ m in distance) by micromanipulation (Fig. 2b, c and Supplementary Fig. 2). The beads,  $\sim 40$   $\mu$ m in diameter, gradually melted on the medium. We found that both TfCRP1 and TfCRP3 showed significant activity to attract pollen tubes in a concentration-dependent manner (Fig. 2d, e). At the optimum concentration,  $52.5 \pm 3.4\%$  (mean  $\pm$  s.d.;  $n = 34$ ; 40 nM) and  $65.7 \pm 7.4\%$  ( $n = 44$ ; 4 nM) of pollen tubes were attracted to TfCRP1 and TfCRP3, respectively. When we used the protruding embryo sac of *T. fournieri* instead of the beads, 67.6% of the pollen tubes were attracted to the embryo sac



**Figure 2 | *In vitro* pollen tube attraction assay using recombinant CRPs.** **a**, TfCRP3 mixed with 10 kD Alexa Fluor (asterisk) was injected in front of a pollen tube through a micropipette at time points of 00:16 (mm:ss) and 10:20. Spectral colours correspond to the intensity of fluorescence (concentration of the Alexa Fluor dye), with white representing the highest level (see colour scales). Scale bar, 20 μm. **b**, **c**, Arrowheads mark the position of the tip of pollen tubes when gelatin beads (asterisks) containing TfCRP1 (**b**) and TfCRP3 (**c**) were placed (00:00). Double arrowheads indicate the tips of the tubes growing towards the beads. Green fluorescence in the assaying of

TfCRP1 shows the remaining concentration gradient of 10 kD Alexa Fluor at 20 min after placement of the bead. Scale bar, 20 μm. **d**, **e**, Pollen tube (PT) attraction activity of TfCRP1 (**d**) and TfCRP3 (**e**), which depends not only on the concentration of peptides but also on the competency, species origin of the pollen tube and preparation of polypeptides (heat treatment, 95 °C for 5 min). The purified and refolded polypeptides were directly used without further fractionation, implying that both active and non-active forms of polypeptides would contribute to the concentration. Data are means and standard deviation ( $n > 3$  with more than 10 pollen tubes per replicate).

( $n = 34$ ). Surprisingly, even a very low amount of TfCRP1 and TfCRP3 (40 pM) already showed activity, which means that only  $\sim 10^3$  molecules in a bead are sufficient to attract pollen tubes. In contrast, TfCRP2 ( $> 2.7 \mu\text{M}$ ) did not show attraction activity ( $7.8 \pm 1.2\%$ ;  $n = 39$ ), at least in our conditions using the chemically synthesized peptide.

In the *in vitro* *Torenia* system, pollen tubes germinated *in vitro* are not competent to respond to the attraction signal from the synergid cells<sup>15</sup>. Pollen tubes are stimulated by unknown sporophytic female factors to acquire their competence<sup>1</sup>. Moreover, pollen tubes of a closely related species, *Lindernia micrantha*, do not respond to the attraction signal derived from the synergid cells of *T. fournieri*<sup>4</sup>. Consistently, TfCRP1 and TfCRP3 did not show any activity on pollen tubes germinated on medium or on tubes of *L. micrantha* (Fig. 2d, e). The latter result was also consistent with the specificity of antibodies against TfCRP1 and TfCRP3, which did not label the filiform apparatus of *L. micrantha* (Supplementary Fig. 1e). Thus, the activity of TfCRP1 and TfCRP3 was similar to that of the attraction signal derived from the synergid cells of *T. fournieri*. We also confirmed that SCR (also known as SP11) of *Brassica*, one of the DEFLs (8 cysteines) that is a pollen determinant of self-incompatibility<sup>16</sup>, showed no activity in the beads assay ( $1.7 \pm 3.7\%$ ;  $n = 57$ ; 40 nM of SP11 (S8-allele)). The empty vector control also showed no activity ( $5.4 \pm 4.9\%$ ;  $n = 35$ ). Hereafter, we named the two CRPs showing the ability to attract the pollen tube as LUREs (TfCRP1 as LURE1 and TfCRP3 as LURE2). LURE1 and LURE2 were heat-labile (Fig. 2d, e). We confirmed using LURE2 that the refolding step was important for its activity ( $20.6 \pm 12.5\%$  attraction without refolding;  $n = 39$ ; 4 nM).

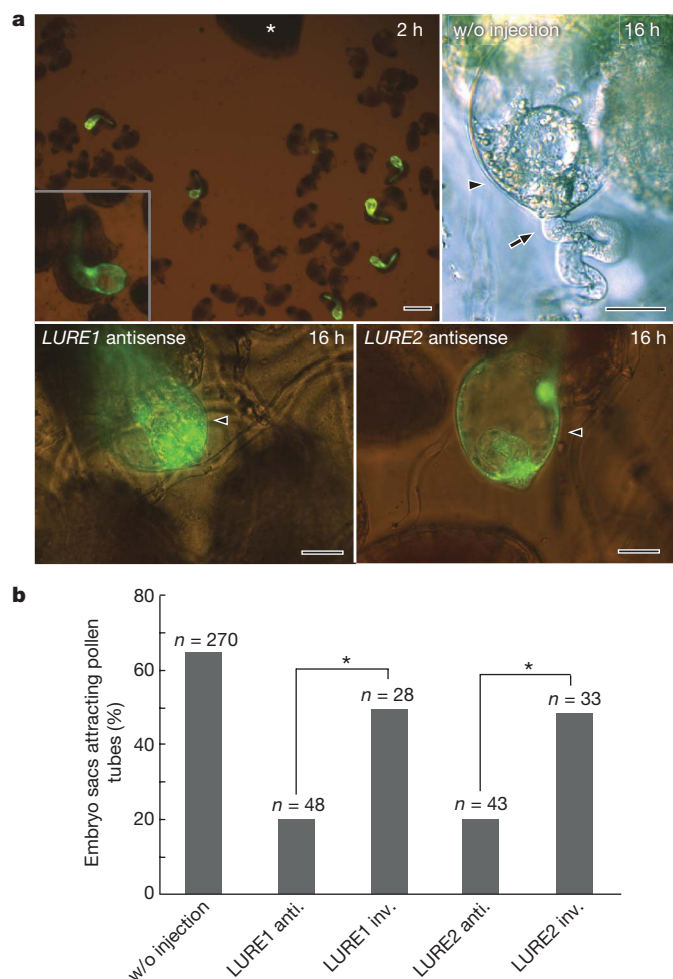
To support further the finding that LUREs are involved in pollen tube attraction by the synergid cell, we tried to downregulate LUREs

by injecting morpholino (MO) antisense oligomers (oligos)<sup>17</sup> into the protruding embryo sac cells of *T. fournieri* (Fig. 3). Efficient microinjection into plant cells with an intact cell wall and high turgor pressure was achieved by developing a laser-assisted thermal-expansion microinjector. We injected fluorescein isothiocyanate (FITC)-labelled MO antisense oligos into the central cell, which is the easiest cell to access. FITC fluorescence spread into the cytosol and nuclei of every gametophytic cell, including the two synergid cells (Fig. 3a), possibly through their plasmodesmata<sup>18</sup>.

At the start of cultivation, we injected a mixture of MO antisense oligos against the translation initiation site, the splice donor site and the splice acceptor site. Because chemically synthesized TfCRP2 did not show pollen tube attraction, we tested antisense oligos for LURE1 (TfCRP1) and LURE2 (TfCRP3). At 16 to 18 h after the start of cultivation, we measured whether pollen tube attraction occurred in each injected embryo sac (Fig. 3). We found that antisense oligos for LURE1 and LURE2 specifically decreased the frequency of attraction, compared to oligos with inverted sequences ( $P = 0.0083$  for LURE1 and  $P = 0.0113$  for LURE2; Fisher's exact test). These results further indicated that LURE1 and LURE2 are involved in pollen tube attraction by the synergid cells.

After a long search for attractants<sup>1</sup>, it is now widely accepted that pollen tubes receive multiple, stage-specific guidance cues before arriving at the range of attraction governed by the synergid cell (a few hundred micrometers at most). Although it has to be elucidated whether other chemoattractants are actually involved in such guidance cues, various candidates of attractants have been reported, including a lily small cell wall protein chemocyanin<sup>19</sup> and water gradient in lipids<sup>20</sup>, as factors for directional growth of pollen tubes at the stigma. GABA ( $\gamma$ -aminobutyric acid) has been suggested as sporophytic signal for ovule targeting<sup>21</sup>. A candidate for the attractant derived from the





**Figure 3 | Microinjection of MO antisense oligos for LUREs.** **a**, MO antisense oligos labelled with FITC (green fluorescence) were injected into the protruding embryo sac of *T. fournieri* at the start of cultivation (2 h). An inset in the top left panel shows an injected ovule. At 16 h after the start of cultivation, ovules without (w/o) injections showed normal pollen tube attraction (arrow) to the protruding embryo sac (arrowhead), whereas many of the injected ovules showed inhibition of pollen tube attraction. Asterisk indicates the cut end of the pollinated pistil. Scale bar in the top left panel is 200  $\mu$ m, and in other panels is 20  $\mu$ m. **b**, Injection of antisense oligos (anti.) showed a reduced frequency of attraction compared to injection of inverted oligos (inv.). The asterisk shows significant differences by Fisher's exact test ( $P < 0.02$ ). Data are frequencies for the total number of embryo sacs examined ( $n$ ) in more than five replicates.

synergid cell has been reported in maize; *Zea mays* EGG APPARATUS-1 (ZmEAI, Q5G8Z3), a small membrane protein predominantly expressed in the synergid cell, is likely to spread to the micropylar nucellar cell walls of the ovule and is necessary for pollen tube guidance<sup>22</sup>. Unlike LUREs, these candidate molecules do not belong to CRP subfamilies.

In contrast, flowering plants use various CRPs at different steps of the pollen–pistil interaction (Supplementary Fig. 3). SCR/SP11, for example, belongs to another subgroup of DEFLs containing eight cysteines and acts as the male determinant of the self-incompatibility reaction in the Brassicaceae<sup>16</sup>. Stigma/stylar cysteine-rich adhesin (SCA) belongs to lipid transfer proteins and is involved in adhesion of the pollen tube to the extracellular matrix of female tissues<sup>23</sup>. Interestingly, SCA enhanced the activity of chemocyanin synergistically<sup>19</sup>, whereas LURE1 and LURE2 could work solely but not synergistically, as examined by mixing them together ( $50.0 \pm 6.3\%$  attraction;  $n = 48$ ; 400 pM LURE1 and 40 pM LURE2). Other subclasses of DEFLs such as a pollen coat protein PCP-A1 (CAA06465; ref. 24) and *Z. mays*

EMBRYO SAC1 (ZmES1, AAK08132)<sup>9</sup> might be involved in intercellular signalling of gametophytic cells. Most of these CRP subclasses were detected in our EST analyses (Supplementary Fig. 3).

It is known that the synergid cell is a key player controlling various fertilization mechanisms<sup>25–28</sup>. A number of synergid-cell-specific or predominant genes have been found recently<sup>3</sup>, including MYB98, ZmEAI and FERONIA, a receptor-like kinase required for pollen tube growth termination<sup>29</sup>. Our results now indicate that LUREs, members of the CRP super gene family, are the pollen tube attractants derived from the synergid cell of *T. fournieri*. LUREs showed abundant expression, which is not the usual case for a signalling molecule; this might be important to keep a gradient of the concentration. The synergid cell of flowering plants, which locates at the entrance of the female gametophyte, expresses many more kinds of CRPs, which may have evolved originally from components of the plant natural immunity system. It will be exciting to find out the role of the other CRPs, which might include factors for defence, pollen tube attraction and other fertilization mechanisms.

## METHODS SUMMARY

**Plant materials.** *Torenia fournieri* cv. 'blue and white' and *Lindernia micrantha* were used<sup>4</sup>.

**EST and expression analysis.** Ovules of *T. fournieri* were treated with an enzyme solution to detach embryo sacs as protoplasts. Synergid cells were collected by using a micropipette under an inverted microscope. After amplification by PCR<sup>30</sup>, cDNAs were cloned and sequenced to generate >4,000 ESTs. EST contigs were assembled by Phred/Phrap software and were manually checked to determine whether they were CRPs<sup>6</sup> and contained a putative signal peptide (SignalP 3.0). Expression of CRP genes was confirmed by RT- and qRT-PCR.

**Immunological analysis.** Proteins of mature ovaries were purified using chloroform and methanol. Pre-immune and anti-LURE affinity purified antibodies (1:1,000 dilution) and peroxidase-conjugated anti-rabbit IgG (1:5,000 dilution; KPL) were used. For immunostaining, purified antibodies (1:1,000 dilution) and Alexa Fluor 488 anti-rabbit IgG (Invitrogen; 1:1,000 dilution) were used.

**Purification of peptides and in vitro attraction assays.** Recombinant LUREs were expressed in *E. coli* and purified by HPLC. TfCRP2 was chemically synthesized (Operon). After dialysis and refolding, each peptide was injected in front of pollen tubes. For bead assays, peptides were embedded in gelatin beads with Alexa Fluor dyes conjugated with 10 kD dextran (Invitrogen). Each bead was picked with a glass needle and placed in front of pollen tubes.

**Microinjection of MO oligos.** A mixture of MO antisense oligos (Supplementary Table 3; Gene Tools LLC) conjugated with FITC was used to inhibit translation and splicing. We developed a laser injector, which has been commercialized along with consumable supplies by Nepa Gene. A glass needle was loaded from the bottom with MO oligos, a hydrophobic ionic liquid, a laser beam absorbent and an ultraviolet curable resin. After the resin was cured, sealing the bottom end, the MO oligos were injected into cells by heat-expanding the laser absorbent with a laser beam.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 28 October 2008; accepted 10 February 2009.

- Higashiyama, T. & Hamamura, Y. Gametophytic pollen tube guidance. *Sex. Plant Reprod.* **21**, 17–26 (2008).
- Higashiyama, T. et al. Pollen tube attraction by the synergid cell. *Science* **293**, 1480–1483 (2001).
- Punwani, J. A. & Drews, G. N. Development and function of the synergid cell. *Sex. Plant Reprod.* **21**, 7–15 (2008).
- Higashiyama, T. et al. Species preferentiality of the pollen tube attractant derived from the synergid cell of *Torenia fournieri*. *Plant Physiol.* **142**, 481–491 (2006).
- Imre, K. & Kristof, Z. Isolation and osmotic relations of developing megagametophytes of *Torenia fournieri*. *Sex. Plant Reprod.* **12**, 152–157 (1999).
- Silverstein, K. A. et al. Small cysteine-rich peptides resembling antimicrobial peptides have been under-predicted in plants. *Plant J.* **51**, 262–280 (2007).
- Punwani, J. A., Rabiger, D. S. & Drews, G. N. MYB98 positively regulates a battery of synergid-expressed genes encoding filiform apparatus-localized proteins. *Plant Cell* **19**, 2557–2568 (2007).
- Jones-Rhoades, M. W., Borevitz, J. O. & Preuss, D. Genome-wide expression profiling of the *Arabidopsis* female gametophyte identifies families of small, secreted proteins. *PLoS Genet.* **3**, 1848–1861 (2007).
- Cordts, S. et al. ZmES genes encode peptides with structural homology to defensins and are specifically expressed in the female gametophyte of maize. *Plant J.* **25**, 103–114 (2001).

10. Yang, H., Kaur, N., Kiriakopolos, S. & McCormick, S. EST generation and analyses towards identifying female gametophyte-specific genes in *Zea mays* L. *Planta* **224**, 1004–1014 (2006).
11. Kasahara, R. D. *et al.* MYB98 is required for pollen tube guidance and synergid cell differentiation in *Arabidopsis*. *Plant Cell* **17**, 2981–2992 (2005).
12. Silverstein, K. A., Graham, M. A., Paape, T. D. & VandenBosch, K. A. Genome organization of more than 300 defensin-like genes in *Arabidopsis*. *Plant Physiol.* **138**, 600–610 (2005).
13. Cornet, B. *et al.* Refined three-dimensional solution structure of insect defensin A. *Structure* **3**, 435–448 (1995).
14. Yount, N. Y. & Yeaman, M. R. Multidimensional signatures in antimicrobial peptides. *Proc. Natl Acad. Sci. USA* **101**, 7363–7368 (2004).
15. Higashiyama, T., Kuroiwa, H., Kawano, S. & Kuroiwa, T. Guidance *in vitro* of the pollen tube to the naked embryo sac of *Torenia fournieri*. *Plant Cell* **10**, 2019–2031 (1998).
16. Takayama, S. *et al.* Direct ligand–receptor complex interaction controls *Brassica* self-incompatibility. *Nature* **413**, 534–538 (2001).
17. Karkare, S. & Bhatnagar, D. Promising nucleic acid analogs and mimics: characteristic features and applications of PNA, LNA, and morpholino. *Appl. Microbiol. Biotechnol.* **71**, 575–586 (2006).
18. Han, Y. Z., Huang, B. Q., Zee, S. Y. & Yuan, M. Symplastic communication between the central cell and the egg apparatus cells in the embryo sac of *Torenia fournieri* Lind. before and during fertilization. *Planta* **211**, 158–162 (2000).
19. Kim, S. *et al.* Chemocyanin, a small basic protein from the lily stigma, induces pollen tube chemotropism. *Proc. Natl Acad. Sci. USA* **100**, 16125–16130 (2003).
20. Wolters-Arts, M., Lush, W. M. & Mariani, C. Lipids are required for directional pollen-tube growth. *Nature* **392**, 818–821 (1998).
21. Palanivelu, R., Brass, L., Edlund, A. F. & Preuss, D. Pollen tube growth and guidance is regulated by *POP2*, an *Arabidopsis* gene that controls GABA levels. *Cell* **114**, 47–59 (2003).
22. Márton, M. L., Cordts, S., Broadhvest, J. & Dresselhaus, T. Micropylar pollen tube guidance by EGG APPARATUS 1 of maize. *Science* **307**, 573–576 (2005).
23. Park, S. Y. *et al.* A lipid transfer-like protein is necessary for lily pollen tube adhesion to an *in vitro* stylar matrix. *Plant Cell* **12**, 151–163 (2000).
24. Doughty, J. *et al.* PCP-A1, a defensin-like *Brassica* pollen coat protein that binds the S locus glycoprotein, is the product of gametophytic gene expression. *Plant Cell* **10**, 1333–1347 (1998).
25. Cheung, A. Y. & Wu, H. M. Pollen tube guidance—right on target. *Science* **293**, 1441–1442 (2001).
26. Higashiyama, T. The synergid cell: attractor and acceptor of the pollen tube for double fertilization. *J. Plant Res.* **115**, 149–160 (2002).
27. Márton, M. L. & Dresselhaus, T. A comparison of early molecular fertilization mechanisms in animals and flowering plants. *Sex. Plant Reprod.* **21**, 37–52 (2008).
28. Berger, F., Hamamura, Y., Ingouff, M. & Higashiyama, T. Double fertilization — caught in the act. *Trends Plant Sci.* **13**, 437–443 (2008).
29. Escobar-Restrepo, J. M. *et al.* The FERONIA receptor-like kinase mediates male–female interactions during pollen tube reception. *Science* **317**, 656–660 (2007).
30. Sprunck, S. *et al.* The transcript composition of egg cells changes significantly following fertilization in wheat (*Triticum aestivum* L.). *Plant J.* **41**, 660–672 (2005).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank T. Suzuki for help with the analysis of the EST sequences, S. Takayama for providing chemically synthesized and refolded SP11 peptides, W. Uchida for checking the diameter of the glass needles for microinjection by scanning electron microscopy and N. Iwata for assistance in preparing plant materials. This work was supported in part by a Grant-in-Aid for Young Scientists (Start-up), Ministry of Education, Culture, Sports, Science, and Technology (MEXT), Japan (20870020 to M.M.K.); a Grant-in-Aid for Creative Scientific Research, MEXT, Japan (18GS0314-01 to N.S.); a grant from Yamada Science Foundation, Japan (to T.H.); a Grant-in-Aid for Scientific Research (B), MEXT, Japan (19370017 to T.H.); a Grant-in-Aid for Scientific Research on Priority Areas (18075004 to T.H.), and PRESTO project, Japan Science and Technology Agency, Japan (to T.H.).

**Author Contributions** S.O., N.S. and T.H. developed methods to purify and assay recombinant LUREs. H. Tsutsui and T.H. developed methods for the microinjection of the MO oligos and performed the RT–PCR analysis. K.S., N.S. and M.M.K. developed immunological methods to detect LUREs. T.H., S.S. and T.D. constructed the cDNA library of the synergid cell of *T. fournieri* and directed the EST analysis. H. Takeuchi, R.Y., R.D.K., Y.H., A.M., D.S., N.K., T.S., K.I., K.O., M.M., H.N. and M.M.K. assembled and analysed the EST sequences. S.N. constructed expression vectors and purified recombinant LUREs. T.K., A.N., T.D., M.M.K., N.S. and T.H. contributed to the experimental design. T.H. directed the project and wrote the paper with input from co-authors.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to T.H. ([higashi@bio.nagoya-u.ac.jp](mailto:higashi@bio.nagoya-u.ac.jp)).

## METHODS

**Plant materials.** We used *Torenia fournieri* cv. 'blue and white' and *Lindernia micrantha* (*Vandellia angustifolia*)<sup>4</sup> plants. Pollen tubes growing through a cut, pollinated style were cultured with ovules for 14–18 h on solid medium<sup>31</sup>. For the *in vitro* assay using gelatin beads, we used pre-conditioned medium instead of cultivation with ovules, which was prepared by cultivating ovules on the placentas from ten ovaries in 1 ml fresh liquid medium overnight (20 h). The culture was covered with hydrated silicone oil (KF-96-100CS; Shin-Etsu) before assaying and microinjection.

**EST and expression analysis.** Each ovary of *Torenia* was excised in 200 µl of enzyme solution (1% cellulase (Worthington), 0.3% macerozyme RS (Yakult), 0.05% pectolyase Y-23 (Yakult), 5 mM calcium nitrate and 0.4 M mannitol, pH 5.8) in a glass-bottom dish. After incubation at 28 °C for 60 min, detached gametophytic protoplasts were collected using a micropipette (Narishige) under an inverted microscope. After collection of the protoplasts, synergid cells were separately collected and washed in 0.4 M mannitol solution containing 5 mM calcium. Synergid cells were transferred to an Eppendorf tube using a microscope and stored at –80 °C. PolyA<sup>+</sup>-RNA was extracted from 25 synergid cells and used for reverse transcription<sup>30</sup>. The cDNA was amplified by long-distance PCR using an optimal number of 30 cycles, as determined by Southern hybridization of 5-µl cDNA aliquots amplified by 18 to 33 PCR cycles each, using a DIG-labelled probe of *Torenia GAPDH*. This PCR probe was obtained using *Torenia* cDNA and the primers TfgAPfw and TfgAPrev. The amplified cDNAs were gel-purified by using a 200-nucleotide cutoff so that degraded RNAs and primer dimers were removed and that mRNAs encoding small peptides were obtained. The PCR products were cloned into the pGEM-T Easy vector (Promega) by TA cloning. Plasmids from randomly picked clones were sequenced from both directions (Agowa). EST sequences were assembled into contigs using Phred/Phrap software with masking of adenine stretches of 30 or more nucleotides. All 256 contigs were manually checked to determine whether they were CRPs<sup>6</sup> and had a putative signal peptide (SignalP 3.0). RT-PCR analysis for each ovular tissue or cell was performed using samples collected under the inverted microscope. For qRT-PCR, gene-specific TaqMan probes were used (Supplementary Table 3).

**Immunological analysis.** Antigen LUREs with or without refolding were prepared as described below and used to immunize rabbits. Immunoglobulin Gs of the pre-immune serum and anti-LUREs were purified by affinity chromatography (HiTrap Protein G HP, GE Healthcare). For immunoblotting analysis, mature ovaries of *T. fournieri* (~100 mg) were collected and frozen in liquid nitrogen. The ovaries were homogenized in buffer (50 mM Tris-HCl, pH 7.5, 50 mM NaCl, 2.5 mM EDTA, 5 mM DTT (or 20 mM TBP) and a protease inhibitor cocktail (Roche)). Proteins were purified by adding chloroform, methanol and distilled water (DW) (sample:chloroform:methanol:DW was 1:1:4:3 v:v:v). After centrifugation, precipitated proteins were washed with methanol and dried. After electrophoresis using 15% or 16.5% SDS polyacrylamide gels and Prestained XL-Ladder Broad marker (Low; Promega), peptides were blotted onto PVDF membranes (Immobilon-P or -PSQ, Millipore). Purified pre-immune and anti-LURE antibodies (1:1,000 dilution) and peroxidase-conjugated anti-rabbit goat IgG (1:5,000 dilution; KPL) were used. For immunostaining, ovules on the placenta were fixed using 4% paraformaldehyde in PBS for 40 min. After treatment with 0.2% Triton X-100 for 5 min and blocking with 3% BSA for 1 h, samples were treated for 2 h with specific antibodies (1:1,000 dilution). The samples were then

treated for 2 h with Alexa-Fluor-488-conjugated anti-rabbit goat IgG (Invitrogen; 1:1,000 dilution). After excision from the placenta, stained ovules were observed under an epi-fluorescence microscope or a CLSM<sup>32</sup>.

**Purification of peptides and *in vitro* attraction assays.** Coding sequences of LUREs lacking N-terminal signal sequences (Supplementary Table 2) were amplified from the cDNA of *T. fournieri* ovaries and cloned into the vector pGEX-6p-1 (GE Healthcare) or pET-28a(+) (Novagen) to fuse glutathione-S-transferase (GST) tag or poly-histidine (His) tag to the N-terminal, respectively. Peptides were expressed in *E. coli* (BL21-CodonPlus; Stratagene) at 37 °C overnight and purified by a glutathione sepharose column 4B (Amersham) or a Ni column (His-Trap FF). The GST tag was removed by a Precision protease (Amersham; five linker amino acids remained). These peptides without a tag were used as recombinant proteins unless otherwise noted. Predicted mature TfCRP2 was chemically synthesized (Operon). The peptides were dialysed (Spectra/Por3 MWCO:3,500 or Spectra/Por6 MWCO:1,000; Spectrum Laboratories) and refolded for 3 days at 4 °C using glutathione (reduced and oxidized forms; Wako) and L-arginine ethyl ester dihydrochloride (Sigma). For *in vitro* assays using micropipettes, each peptide in buffer (50 mM Tris-HCl, pH 8.0) was loaded into a glass needle as described below. The microinjector was positioned using a micro-manipulator (Narishige) attached to an inverted fluorescence microscope. Each peptide was injected in front of the pollen tubes. Behaviour of pollen tubes was photographed by a CCD camera or by a time-lapse digital video system (Sigma Koki). For *in vitro* assays using gelatin beads, 1 µl of peptides with His tag was mixed with 2 µl of 10% (w/v) gelatin (Nacalai) pre-melted at 50 °C and 1 µl of 1 mM Alexa Fluor 488, 546 or 568 conjugated with 10 kD dextran (Invitrogen). After adding 150 µl of hydrated silicone oil, they were mixed by vortexing and cooled on ice for formation of gelatin beads. Each bead (approximately 40 µm in diameter) was picked using a glass needle and placed in front of the pollen tubes. The criteria for judgment of 'attracted' and 'non-attracted' pollen tubes are summarized in Supplementary Fig. 2.

**Microinjection of MO oligos.** Three kinds of MO antisense oligos conjugated with FITC (Gene Tools LLC) were designed for inhibition of the translation initiation site and the splice donor and acceptor sites of each LURE. MO oligos with inverted sequences were also designed as negative controls. After dissolution in distilled water at a concentration of 1 mM, each 1.5 µl of the three antisense MO oligos, as well as the three inverted MO oligos, were mixed (each 0.33 mM) and centrifuged at 30,000g for 20 min to precipitate impurities. We developed and commercialized a method of laser-assisted microinjection (Nepa Gene). A glass needle, with an outer diameter at the tip of 0.2–0.3 µm, was made by pulling glass capillaries (GDC1; Narishige) with P-97 (Sutter) or MCF-100 (Nepa Gene) micropipette pullers. The needle was loaded with MO oligos, a hydrophobic ionic liquid, a laser beam absorbent and an ultraviolet curable resin through the bottom. After the resin was cured by irradiation with ultraviolet light to seal the bottom end, the needle was placed in the holder of the laser injector (LTM-1000; Nepa Gene). MO oligos were injected into cells by heating and expanding the laser absorbent by use of a laser beam. The amount of injected oligonucleotides was adjusted to ~1% (v/v) of the volume of the central cell by observing the fluorescence of FITC with a CCD camera. Micro-injection was performed during the first 3 h after the start of cultivation.

31. Higashiyama, T. & Inatsugi, R. Comparative analyses of biological models used in the study of pollen tube growth. *Plant Cell Monogr.* 3, 265–286 (2006).
32. Ingouff, M. *et al.* Distinct dynamics of HISTONE3 variants between the two fertilization products in plants. *Curr. Biol.* 17, 1032–1037 (2007).



## LETTERS

# The DNA-encoded nucleosome organization of a eukaryotic genome

Noam Kaplan<sup>1\*</sup>, Irene K. Moore<sup>3\*</sup>, Yvonne Fondufe-Mittendorf<sup>3</sup>, Andrea J. Gossett<sup>4</sup>, Desiree Tillo<sup>5</sup>, Yair Field<sup>1</sup>, Emily M. LeProust<sup>6</sup>, Timothy R. Hughes<sup>5,7,8</sup>, Jason D. Lieb<sup>4</sup>, Jonathan Widom<sup>3</sup> & Eran Segal<sup>1,2</sup>

Nucleosome organization is critical for gene regulation<sup>1</sup>. In living cells this organization is determined by multiple factors, including the action of chromatin remodellers<sup>2</sup>, competition with site-specific DNA-binding proteins<sup>3</sup>, and the DNA sequence preferences of the nucleosomes themselves<sup>4–8</sup>. However, it has been difficult to estimate the relative importance of each of these mechanisms *in vivo*<sup>7,9–11</sup>, because *in vivo* nucleosome maps reflect the combined action of all influencing factors. Here we determine the importance of nucleosome DNA sequence preferences experimentally by measuring the genome-wide occupancy of nucleosomes assembled on purified yeast genomic DNA. The resulting map, in which nucleosome occupancy is governed only by the intrinsic sequence preferences of nucleosomes, is similar to *in vivo* nucleosome maps generated in three different growth conditions. *In vitro*, nucleosome depletion is evident at many transcription factor binding sites and around gene start and end sites, indicating that nucleosome depletion at these sites *in vivo* is partly encoded in the genome. We confirm these results with a micrococcal nuclease-independent experiment that measures the relative affinity of nucleosomes for ~40,000 double-stranded 150-base-pair oligonucleotides. Using our *in vitro* data, we devise a computational model of nucleosome sequence preferences that is significantly correlated with *in vivo* nucleosome occupancy in *Caenorhabditis elegans*. Our results indicate that the intrinsic DNA sequence preferences of nucleosomes have a central role in determining the organization of nucleosomes *in vivo*.

We sought to establish the extent to which the DNA sequence determines nucleosome organization in living cells. Our strategy, previously used by others for two yeast promoters<sup>12</sup>, was to compare *in vivo* nucleosome organization with that obtained by an *in vitro* assembly procedure using only purified nucleosomes and purified DNA. To obtain a genome-wide map of nucleosome occupancy governed solely by nucleosome sequence preferences, we purified chicken erythrocyte histone octamers and assembled them on purified yeast genomic DNA by salt gradient dialysis<sup>13</sup>. We then isolated mononucleosomes by micrococcal nuclease digestion, and used parallel sequencing to determine nucleosome positions. We performed two independent experiments, resulting in ~10,000,000 DNA sequence reads that map uniquely to the yeast genome. For comparison to *in vivo* nucleosome positions, we isolated mononucleosomes from living cells<sup>5,7,9,10</sup>, and obtained ~25,000,000 sequence reads from 6 independent experiments. For each map, we determined the average nucleosome occupancy at every base pair, calculated as the log-ratio between the number of reads that cover that base

pair and the genome-wide average coverage per base pair (see Methods).

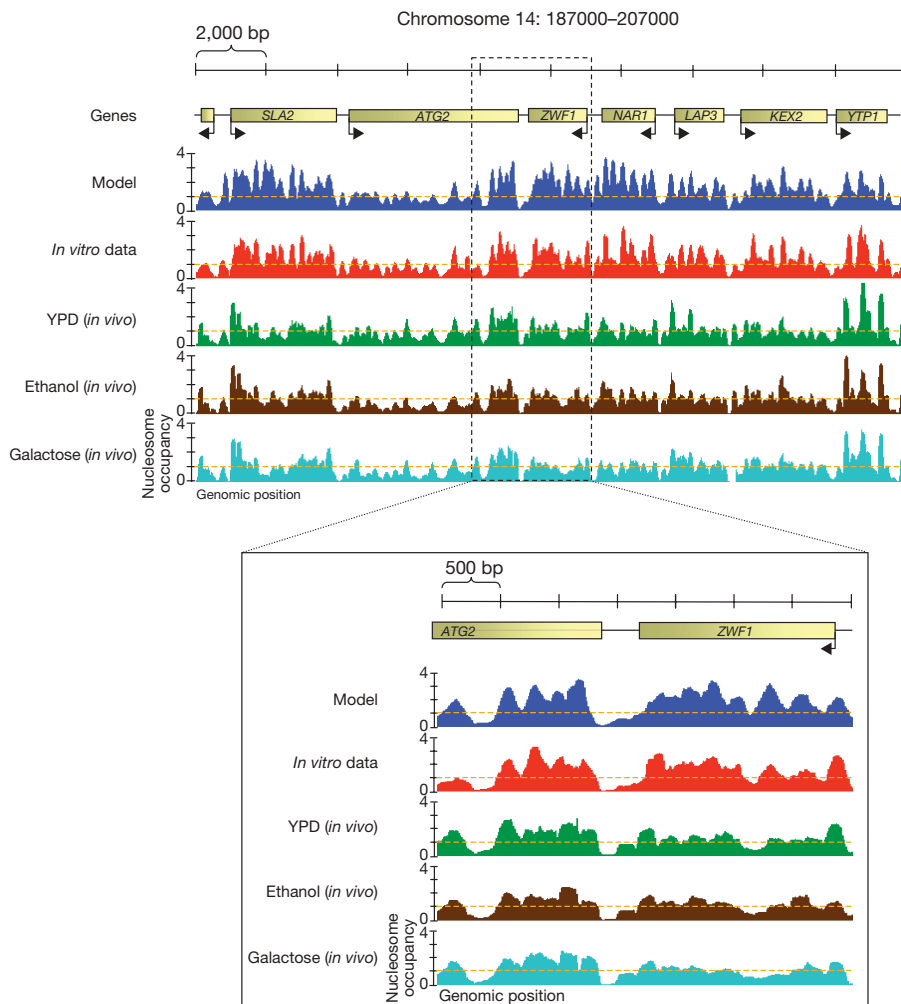
The nucleosome organizations of the *in vitro* and *in vivo* maps are notably similar, although not identical (Fig. 1), with a correlation of 0.74 between the nucleosome occupancy per base pair (Fig. 2a). On the scale of individual nucleosomes, the *in vitro* data separate regions that are enriched in nucleosomes *in vivo* from regions depleted of nucleosomes with high accuracy (Supplementary Fig. 1). Similarly, we found a significant correspondence between the positions of stable nucleosomes in the two maps (Supplementary Fig. 2). This high degree of similarity between the maps indicates that nucleosome sequence preferences have a dominant role in determining *in vivo* nucleosome organization.

The correlation between the maps is not uniform across the genome. We found a higher correlation between the maps at non-promoter intergenic regions located at ends of convergently transcribed genes (0.83) and a lower correlation at promoter (0.69) and coding (0.69) regions. In addition, the depletion level *in vivo* relative to that measured *in vitro* at coding regions increases with the expression level of the associated genes (Fig. 2b). These results indicate that transcription factors, chromatin regulators and active transcription influence the resulting nucleosome organization *in vivo*.

Because the nucleosome organization *in vitro* is determined only by the DNA sequence, we asked whether we could derive rules that are predictive of nucleosome positioning and occupancy. For each of the 1,024 sequences of length 5 base pairs, we computed the average nucleosome occupancy of that sequence across all of its instances in the genome. We found a near perfect agreement (correlation of 0.98) between the average occupancy of these 5-base-pair sequences *in vivo* and *in vitro* (Fig. 3a). Many 5-base-pair sequences showed strong preferences for nucleosome-enriched or nucleosome-depleted regions. For example, AAAAA has the lowest average nucleosome occupancy both *in vivo* and *in vitro*, consistent with the reduced nucleosome affinity that poly(dA-dT) sequences have *in vitro*<sup>14</sup>, and with the nucleosome depletion observed over poly(dA-dT) sequences *in vivo*<sup>9,15</sup>. Consistent with previous reports<sup>4,5,11,16</sup>, we also found clear ~10-bp periodicities of dinucleotides along the nucleosome length, both *in vitro* and *in vivo* (Fig. 3b, c). Notably, the dynamic range of these periodicities is greater *in vitro*, suggesting that the fraction of nucleosomes positioned by these periodic motifs *in vitro* is greater than that *in vivo*. This difference may be due to the action of chromatin remodellers and transcription factors *in vivo*, which may cause nucleosomes to deviate from the locations dictated by the nucleosome sequence preferences. The higher

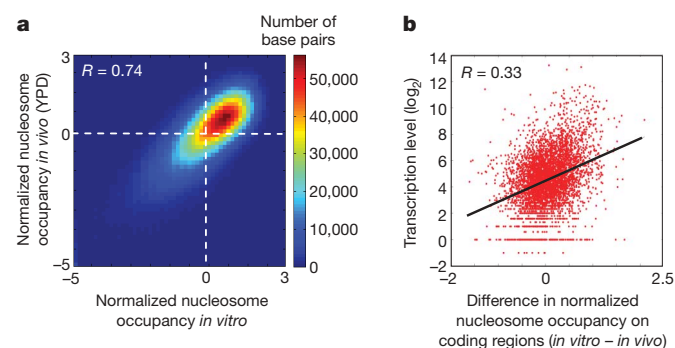
<sup>1</sup>Department of Computer Science and Applied Mathematics, <sup>2</sup>Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot 76100, Israel. <sup>3</sup>Department of Biochemistry, Molecular Biology, and Cell Biology, Northwestern University, 2153 Sheridan Road, Evanston, Illinois 60208, USA. <sup>4</sup>Department of Biology, Carolina Center for Genome Sciences, and Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. <sup>5</sup>Department of Molecular Genetics, University of Toronto, Toronto, Ontario M5S 1A8, Canada. <sup>6</sup>Agilent Technologies Inc., Genomics—LSSU, 5301 Stevens Creek Boulevard, MS 3L/MT Santa Clara, California 95051, USA. <sup>7</sup>Terrence Donnelly Centre for Cellular & Biomolecular Research, <sup>8</sup>Banting and Best Department of Medical Research, 160 College Street, Toronto, Ontario M5S 3E1, Canada.

\*These authors contributed equally to this work.



**Figure 1 | The intrinsic DNA-encoded nucleosome organization at a typical genomic region.** Shown are the four different maps of nucleosome occupancy measured in this study for a typical 20,000-bp-long genomic region: the *in vitro* map, which reflects only the intrinsic nucleosome sequence preferences, and *in vivo* yeast maps for three different growth conditions (YPD, ethanol and galactose). Each track plots the measured nucleosome occupancy per base pair, computed by summing all of the nucleosome reads obtained in that experiment, and dividing that number by the average number of reads per base pair across the genome. The line of  $y = 1$  thus represents the genome-wide average and is shown as a dashed orange line. The average nucleosome occupancy predictions from our model are shown in blue.

concentration of nucleosomes *in vivo* relative to the concentration used to create our *in vitro* map may also contribute to this difference, because higher nucleosome concentrations generally increase the



**Figure 2 | *In vitro* and *in vivo* maps are highly similar.** **a**, Shown is a density dot plot comparison of the normalized nucleosome occupancy per base pair in the *in vitro* (x axis) and *in vivo* (y axis) maps (see Methods). Values above zero indicate nucleosome enrichment relative to the genome-wide average. The colour of each point represents the number of base pairs that map to that point in the graph. The Pearson correlation between the maps is indicated. **b**, Nucleosome depletion *in vivo* relative to *in vitro* over coding regions increases with the expression level of associated genes. Shown is a dot plot comparison between the expression level of every yeast gene (measured in ref. 26) and the difference between the average normalized nucleosome occupancy of the coding region of that gene in the *in vitro* map compared with the *in vivo* map (that is, higher values indicate larger nucleosome depletion *in vivo* relative to *in vitro*). The Pearson correlation of the dot plot is indicated.

contribution of non-specific binding, thus diminishing the contribution of the  $\sim 10$ -bp sequence periodicities. Nevertheless, the conservation of the  $\sim 10$ -bp dinucleotide periodicities and the near-identity of 5-base-pair nucleosome occupancies demonstrate that nucleosomes have clear sequence preferences that are highly similar *in vitro* and *in vivo*.

To test whether general sequence-based rules can be derived from our *in vitro* data and be used to predict nucleosome occupancy *in vivo*, we constructed a simple probabilistic model based on both the global preferences over sequences of length 5 and the position-dependent dinucleotide preferences<sup>5,17</sup>, which scores the nucleosome formation potential of every 147-bp sequence. Importantly, this model is learned only from *in vitro* nucleosome data, and therefore represents only nucleosome sequence preferences, whereas previous models<sup>5–8,18</sup>, which were learned from *in vivo* data, may also capture sequence preferences of other factors<sup>7</sup>, as well as indirect effects due to chromatin remodelling activities. We tested the model in a cross-validation scheme in which the nucleosome occupancy of each chromosome was predicted using a model that was constructed from the data from all other chromosomes. Our model has high correlations of 0.89 and 0.75 with the *in vitro* and *in vivo* maps, respectively (Fig. 3d, e), and separates nucleosome-enriched regions from nucleosome-depleted regions (Supplementary Fig. 3), indicating that the model successfully identified general predictive rules for the sequence preferences of nucleosomes.

If nucleosome sequence preferences are important in other eukaryotes, then our model should also be predictive of their *in vivo* nucleosome organization. Indeed, we found a good (0.60) correlation between the nucleosome occupancy per base pair predicted by our

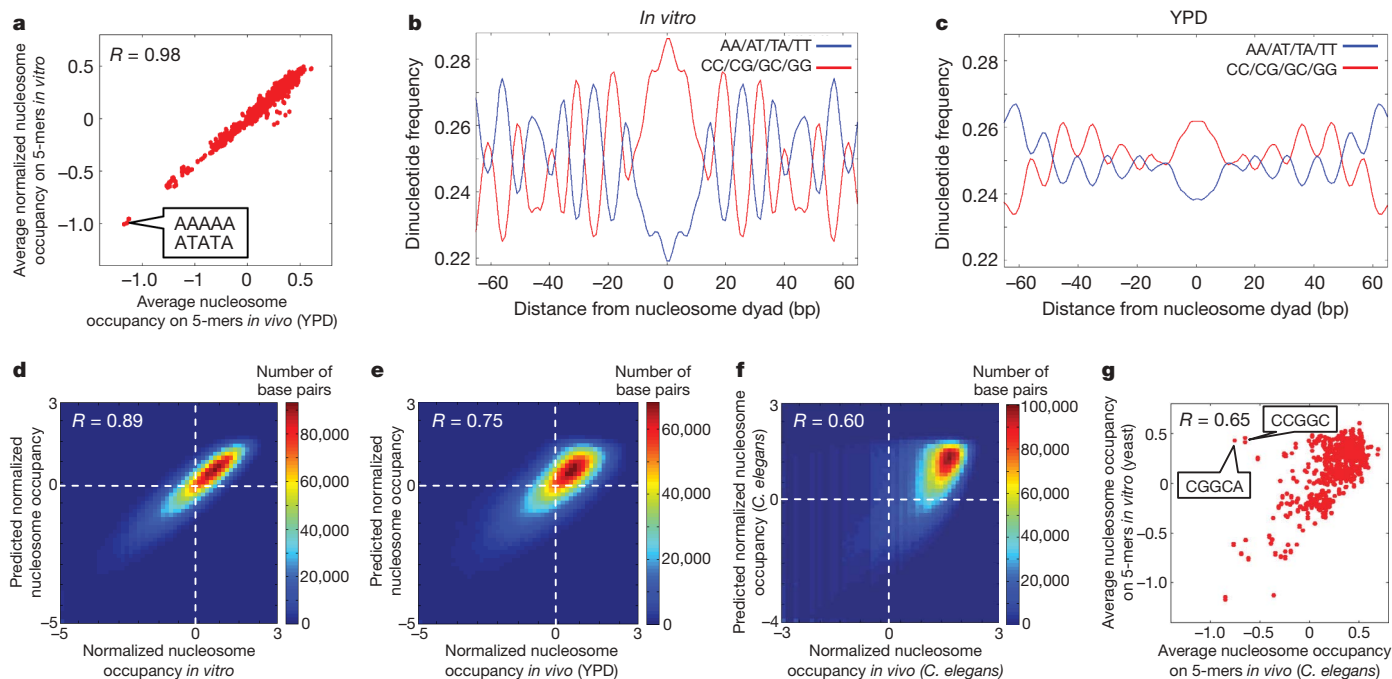
*in vitro* yeast-based model and that of *in vivo* nucleosome occupancy in *C. elegans*<sup>19</sup> (Fig. 3f). Moreover, our model classifies nucleosome-enriched regions from nucleosome-depleted regions in *C. elegans* with high accuracy (Supplementary Fig. 4), and the 5-base-pair sequence preferences of the *C. elegans in vivo* map agree well with those of the yeast *in vitro* map (Fig. 3g). The poorer classification performance in comparison with yeast may indicate that factors other than the DNA sequence preferences make a greater contribution to nucleosome organization in more complex eukaryotes. Alternatively, the poorer performance may indicate that distinct sequence types are present in *C. elegans* for which our yeast *in vitro* data do not provide statistics. Nonetheless, our model is significantly correlated with the *in vivo* nucleosome organization across *C. elegans*.

We next compared the DNA-encoded nucleosome organization of the *in vitro* map with nucleosome organization under growth conditions that cause substantial transcriptional changes relative to log-phase growth in rich medium (that is glucose). In addition to our map obtained from yeast cells grown in rich medium, we also measured the nucleosome organization of yeast cells grown separately in galactose, and in ethanol, and found that the overall nucleosome occupancy is very similar between all three *in vivo* maps, although localized differences are apparent (Fig. 1 and Supplementary Fig. 5). All three *in vivo* maps are highly correlated with the *in vitro* map and show the sequence characteristics seen *in vitro* (Supplementary Fig. 6). These results imply that intrinsic sequence preferences of nucleosomes have a dominant role in determining nucleosome

organization in several growth conditions, with local, condition-specific changes superimposed.

To address concerns regarding biases that may be caused by the sequence specificity of micrococcal nuclease<sup>20</sup> and possible biases in parallel sequencing, we performed a different kind of *in vitro* experiment that measures the relative nucleosome affinity of ~40,000 double-stranded 150-bp oligonucleotides without the use of micrococcal nuclease or parallel sequencing. The resulting 5-base-pair nucleosome sequence preferences are in excellent agreement with those discovered in the genome-wide *in vitro* reconstitution (correlation of 0.83), and there is a good correlation (0.51) between the measured oligonucleotide affinities and those predicted by the model constructed from the genome-wide *in vitro* map (Supplementary Fig. 7). These results are wholly independent of either micrococcal nuclease or parallel sequencing, and thus confirm that the sequence specificities derived from our previous experiments were caused by intrinsic nucleosome preferences, rather than being an artefact of our experimental approach.

Previous studies identified nucleosome depletion around transcription start and stop sites<sup>5–7,9–11</sup>. However, because these studies were based on *in vivo* data, it was not possible to determine which mechanism accounted for the observed patterns. The *in vitro* and *in vivo* maps show highly similar stereotypic nucleosome depletion at translation end sites, indicating that this depletion is largely encoded by nucleosome sequence preferences (Fig. 4b and Supplementary Fig. 8). The two maps also show stereotypic nucleosome depletion



**Figure 3 | The *in vitro* sequence preferences of nucleosomes are highly similar to those of nucleosome-bound sequences *in vivo* and are predictive of nucleosome occupancy in *C. elegans*.** **a**, Comparison of genome-wide relative nucleosome occupancy of nucleosomes over sequences of length 5. For the *in vitro* and *in vivo* maps of nucleosome occupancy, we separately computed the average normalized nucleosome occupancy of each of the 1,024 sequences of length 5, across all of its instances in the genome. Shown is a comparison between the distributions of these 5-base-pair sequences in both maps. Also shown is the Pearson correlation between these distributions. **b**, Position-dependent sequence preferences of nucleosomes in the *in vitro* map. We aligned the individual nucleosome reads in the *in vitro* nucleosome collection. Shown is the fraction (3-bp moving average) of AA/AT/TT/TA and CC/CG/GC/GG dinucleotides at each position of the alignment. **c**, Same as **b**, for the *in vivo* map. **d**, Shown is a density dot plot comparison between the normalized nucleosome occupancy per base pair in the *in vitro* map (*x* axis) and the normalized nucleosome occupancy per base

pair predicted by our cross-validated computational model of nucleosome sequence preferences (*y* axis). Values above zero indicate nucleosome enrichment relative to the genome-wide average. The colour of each point represents the number of base pairs that map to that point in the graph. The Pearson correlation between the maps is indicated. **e**, Same as **d**, comparing our model predictions to the *in vivo* map. **f**, *In vitro* nucleosome sequence preferences on yeast genomic DNA are predictive of the *in vivo* nucleosome organization in *C. elegans*. Same as **d**, comparing our model predictions and the *in vivo* nucleosome occupancy map of *C. elegans* on chromosome 2 (ref. 19). **g**, Comparison of yeast nucleosome sequence preferences *in vitro* and those of *C. elegans in vivo*. For each of the maps we separately computed the average normalized nucleosome occupancy of every possible sequence of length 5. For *C. elegans*, we performed these computations on chromosome 2. Shown is a comparison of these 5-base-pair sequence distributions between the yeast *in vitro* map and the *in vivo* map of *C. elegans*, along with the Pearson correlation between these distributions.

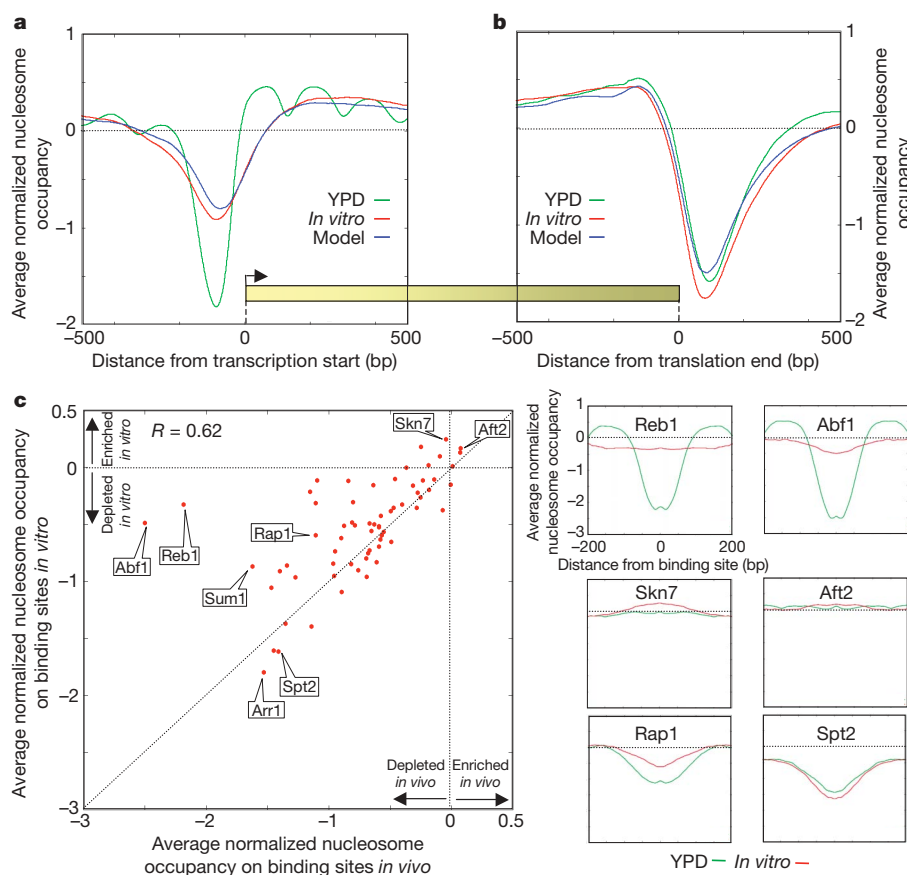


at transcription start sites, indicating that this depletion is also partly encoded by nucleosome sequence preferences (Fig. 4a). However, the level of depletion around transcription start sites *in vitro* is smaller than *in vivo*, indicating that transcription factors, chromatin remodellers, the transcription initiation machinery and other mechanisms also contribute to the depletion. Another difference around start sites (Fig. 4a) is the longer-range ordering of nucleosomes into coding regions observed in the *in vivo* map. This may be partly explained by the higher nucleosome concentration *in vivo* (see Methods), which causes increased ordering by statistical positioning<sup>21</sup>.

Nucleosome depletion has also been reported around transcription factor binding sites *in vivo*<sup>7,9,11</sup>. In both our *in vivo* and *in vitro* maps, nucleosome depletion was observed, on average, around the chromatin-immunoprecipitation-determined binding sites<sup>22</sup> of most transcription factors. The maps also agree on the degree of depletion around binding sites (correlation of 0.62; Fig. 4c and Supplementary Fig. 9). These results indicate that nucleosome depletion around regulatory factor binding sites is partly encoded in the genome's intrinsic nucleosome organization, and that this intrinsic organization may facilitate transcription initiation and assist in directing transcription factors to their appropriate sites in the genome<sup>5,23</sup>.

Binding sites for the yeast factors Abf1 and Reb1 show the largest deviation from the above agreement, with sites for both factors being more depleted *in vivo* than *in vitro*. Notably, both factors are highly abundant and influence chromatin structure<sup>24,25</sup>, indicating that the depletion around their sites *in vivo* may be attributable to their own action. The large nucleosome depletion over these factor sites resulted in them being major components of a nucleosome occupancy model created from *in vivo* data<sup>7</sup>. However, because the model constructed here is based solely on *in vitro* data, in which these sites have only moderate depletion, they are not major components of our model.

In summary, we find marked similarities between the nucleosome organization governed only by the DNA sequence preferences of nucleosomes, and the organization of nucleosomes *in vivo* measured under different growth conditions. This result indicates that nucleosome sequence preferences are important determinants of nucleosome organizations *in vivo*. Our analysis indicates that genomes may use their intrinsically encoded nucleosome organization to facilitate functions such as transcription factor binding and transcription. Despite the overall similarity between the *in vitro* and *in vivo* maps, there are differences, consistent with previous studies showing that



**Figure 4 | The intrinsic nucleosome organization over transcripts and transcription factor binding sites.** **a**, For the *in vitro* and *in vivo* nucleosome occupancy maps, and for our model, shown is the normalized nucleosome occupancy per base pair around the transcription start site, averaged across all yeast genes. The long-range ordering of nucleosome occupancy which is present in the *in vivo* maps but not in the *in vitro* map may be partly explained by the lower nucleosome concentration in which the *in vitro* experiment was carried out (see Methods), because higher nucleosome concentration *in vivo* is predicted to cause long-range ordering of nucleosome arrays<sup>21</sup>. **b**, Same as **a**, but around translation end sites of genes (translation end was chosen because transcription end sites are poorly annotated). The depletion around gene ends may be due to the presence of termination signals<sup>27</sup>, which disfavour nucleosome formation *in vitro* (Supplementary Fig. 8). The fact that these signals tend to occur in a specific

orientation with respect to the direction of transcription<sup>27</sup> is consistent with a function in transcript processing, but does not exclude the possibility that one or more of these motifs functions primarily to disfavour nucleosomes. **c**, Comparison of the nucleosome occupancy over transcription factor binding sites between the *in vitro* and the YPD *in vivo* maps. For each transcription factor with at least 50 functional binding sites<sup>22</sup>, we computed, separately for the *in vivo* and *in vitro* maps, the average normalized nucleosome occupancy over its binding sites. Shown is a comparison of these nucleosome occupancies per factor, between the *in vivo* and *in vitro* maps, along with the Pearson correlation between them. For six factors taken from different regions of the plot, we also show the average normalized nucleosome occupancy around those factors' binding sites, for both the *in vitro* and the *in vivo* maps.

factors other than nucleosome sequence preferences contribute to nucleosome organization *in vivo*. Future studies will focus on understanding how nucleosomes are remodelled locally and the function of such remodelling in transcriptional regulation.

## METHODS SUMMARY

*In vivo* maps from yeast nucleosome DNAs were prepared from log-phase cells grown in rich medium (YPD, six independent replicates) as described previously<sup>5</sup>, as well as from cells grown in YP media supplemented with 2% galactose (three replicates) or 2.8% ethanol (four replicates) instead of glucose. The resulting DNAs were subjected to Illumina sequencing-by-synthesis. For the *in vitro* map, histone octamer was purified from chicken erythrocytes, assembled on purified yeast genomic DNA by salt gradient dialysis<sup>13</sup>, digested with micrococcal nuclease and subjected to Illumina sequencing (two independent replicates). The resulting *in vitro* map has a lower concentration of nucleosomes along the DNA than obtained *in vivo*. This technical limitation was necessitated by our finding that reconstitutions at the *in vivo* stoichiometry on long genomic DNA resulted in insoluble chromatin that was inaccessible to micrococcal nuclease. We mapped the resulting reads to the genome and removed reads that mapped to multiple genomic locations. We extended the nucleosome reads of each experiment to the average nucleosome length in that experiment (always between 140–170 bp). For each map, we then calculated the normalized nucleosome occupancy at every base pair as the log-ratio between the number of reads that cover that base pair and the average number of reads per base pair across the genome. We then set the genomic mean in each sample to zero by subtracting the genome-wide mean from every base pair. The independent replicates for each experiment type were in excellent agreement, so we averaged the replicates within each type. The resulting tracks are termed normalized nucleosome occupancy throughout the manuscript. The detailed formulation of our sequence-based model for nucleosome positioning is given in the Methods and is similar to that described in ref. 17, except that it was learned using only the *in vitro* data. For our data, results and model, see <http://genie.weizmann.ac.il/pubs/nucleosomes08/>, and GEO accession number GSE13622.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 2 October; accepted 26 November 2008.**

**Published online 17 December 2008.**

- Kornberg, R. D. & Lorch, Y. Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell* **98**, 285–294 (1999).
- Vignali, M., Hassan, A. H., Neely, K. E. & Workman, J. L. ATP-dependent chromatin-remodeling complexes. *Mol. Cell. Biol.* **20**, 1899–1910 (2000).
- Korber, P., Luckenbach, T., Blaschke, D. & Horz, W. Evidence for histone eviction in trans upon induction of the yeast PHO5 promoter. *Mol. Cell. Biol.* **24**, 10965–10974 (2004).
- Satchwell, S. C., Drew, H. R. & Travers, A. A. Sequence periodicities in chicken nucleosome core DNA. *J. Mol. Biol.* **191**, 659–675 (1986).
- Segal, E. *et al.* A genomic code for nucleosome positioning. *Nature* **442**, 772–778 (2006).
- Ioshikhes, I. P., Albert, I., Zanton, S. J. & Pugh, B. F. Nucleosome positions predicted through comparative genomics. *Nature Genet.* **38**, 1210–1215 (2006).
- Lee, W. *et al.* A high-resolution atlas of nucleosome occupancy in yeast. *Nature Genet.* **39**, 1235–1244 (2007).
- Yuan, G. C. & Liu, J. S. Genomic sequence is highly predictive of local nucleosome depletion. *PLoS Comput. Biol.* **4**, e13 (2008).
- Yuan, G. C. *et al.* Genome-scale identification of nucleosome positions in *S. cerevisiae*. *Science* **309**, 626–630 (2005).
- Shivaswamy, S. *et al.* Dynamic remodeling of individual nucleosomes across a eukaryotic genome in response to transcriptional perturbation. *PLoS Biol.* **6**, e65 (2008).

- Albert, I. *et al.* Translational and rotational settings of H2A.Z nucleosomes across the *Saccharomyces cerevisiae* genome. *Nature* **446**, 572–576 (2007).
- Sekinger, E. A., Moqtaderi, Z. & Struhl, K. Intrinsic histone–DNA interactions and low nucleosome density are important for preferential accessibility of promoter regions in yeast. *Mol. Cell* **18**, 735–748 (2005).
- Thastrom, A., Bingham, L. M. & Widom, J. Nucleosomal locations of dominant DNA sequence motifs for histone–DNA interactions and nucleosome positioning. *J. Mol. Biol.* **338**, 695–709 (2004).
- Anderson, J. D. & Widom, J. Poly(dA–dT) promoter elements increase the equilibrium accessibility of nucleosomal DNA target sites. *Mol. Cell. Biol.* **21**, 3830–3839 (2001).
- Iyer, V. & Struhl, K. Poly(dA:dT), a ubiquitous promoter element that stimulates transcription via its intrinsic DNA structure. *EMBO J.* **14**, 2570–2579 (1995).
- Ioshikhes, I., Bolshoy, A., Derenshteyn, K., Borodovsky, M. & Trifonov, E. N. Nucleosome DNA sequence pattern revealed by multiple alignment of experimentally mapped sequences. *J. Mol. Biol.* **262**, 129–139 (1996).
- Field, Y. *et al.* Distinct modes of regulation by chromatin encoded through nucleosome positioning signals. *PLoS Comput. Biol.* **4**, e1000216 (2008).
- Peckham, H. E. *et al.* Nucleosome positioning signals in genomic DNA. *Genome Res.* **17**, 1170–1177 (2007).
- Valouev, A. *et al.* A high-resolution, nucleosome position map of *C. elegans* reveals a lack of universal sequence-dictated positioning. *Genome Res.* **18**, 1051–1063 (2008).
- Horz, W. & Altenburger, W. Sequence specific cleavage of DNA by micrococcal nuclease. *Nucleic Acids Res.* **9**, 2643–2658 (1981).
- Kornberg, R. D. & Stryer, L. Statistical distributions of nucleosomes: nonrandom locations by a stochastic mechanism. *Nucleic Acids Res.* **16**, 6677–6690 (1988).
- MacIsaac, K. D. *et al.* An improved map of conserved regulatory sites for *Saccharomyces cerevisiae*. *BMC Bioinformatics* **7**, 113 (2006).
- Liu, X., Lee, C. K., Granek, J. A., Clarke, N. D. & Lieb, J. D. Whole-genome comparison of Leu3 binding *in vitro* and *in vivo* reveals the importance of nucleosome occupancy in target site selection. *Genome Res.* **16**, 1517–1528 (2006).
- Lascaris, R. F., Groot, E., Hoen, P. B., Mager, W. H. & Planta, R. J. Different roles for Abf1p and a T-rich promoter element in nucleosome organization of the yeast *RPS28A* gene. *Nucleic Acids Res.* **28**, 1390–1396 (2000).
- Raisner, R. M. *et al.* Histone variant H2A.Z marks the 5' ends of both active and inactive genes in euchromatin. *Cell* **123**, 233–248 (2005).
- Nagalakshmi, U. *et al.* The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* **320**, 1344–1349 (2008).
- Graber, J. H., Cantor, C. R., Mohr, S. C. & Smith, T. F. Genomic detection of new yeast pre-mRNA 3'-end-processing signals. *Nucleic Acids Res.* **27**, 888–894 (1999).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank H. Kelkar for bioinformatics support in handling the sequencing files, P. Mieczkowski and J. McPherson for Illumina sequencing (library preparation and instrument operation), and the members of our respective laboratories for technical assistance, discussions and comments on the manuscript. The UNC sequencing facility is funded by the Lineberger Comprehensive Cancer Center and University Cancer Research Funds. J.W. acknowledges the use of instruments at Northwestern University's Keck Biophysics Facility. This work was supported by a grant from CIHR to T.R.H. and C. Nislow, a grant from the NIH to J.D.L., a grant from the NIH to J.W., and grants from the European Research Council (ERC) and NIH to E.S. D.T. holds an NSERC postgraduate scholarship. N.K. is a Clore scholar. E.S. is the incumbent of the Soretta and Henry Shapiro career development chair.

**Author Contributions** This work was performed jointly in the Hughes, Lieb, Widom and Segal laboratories and with substantial contributions from each.

**Author Information** This has been submitted to GEO under accession number GSE13622. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to J.W. (j-widom@northwestern.edu) or E.S. (eran.segal@weizmann.ac.il).

## METHODS

**Parallel sequencing of yeast nucleosomes *in vivo*.** Four samples of yeast nucleosome DNAs ('*in vivo* nucleosomes') were prepared from log-phase cells grown in rich medium (YPD) as described previously<sup>5</sup>. Additionally, samples were prepared from cells grown in YP media supplemented with 2% galactose (two samples) or 2.8% ethanol (two samples) instead of glucose. Five additional samples (two in YPD, one in galactose and two in ethanol) were prepared from cells that were formaldehyde-crosslinked before isolating nucleosomes. Crosslinking was achieved with 1% formaldehyde for 15 min at 30 °C and stopped with 125 mM glycine. The cells were then washed, made into spheroplasts with lyticase as above, treated with micrococcal nuclease (MNase) as described previously<sup>12</sup>, and nucleosomal DNAs isolated as described<sup>10</sup>. All of the resulting DNAs were subjected to Illumina/Solexa sequencing-by-synthesis. The results from crosslinked or uncrosslinked cells for a given growth medium were very similar; therefore, we averaged the replicates within each growth medium (six YPD, four YPetOH, three YPGal).

**Parallel sequencing of nucleosomes reconstituted *in vitro* on yeast genomic DNA.** *S. cerevisiae* genomic DNA was purified from strain YLC8 (*MATa ura3Δ leu2Δ his3Δ met15Δ*) using standard methods with additional steps to remove contaminating RNA (see Supplementary Information). Histone octamer was purified from chicken erythrocytes using salt extraction and hydroxyapatite column chromatography, as described previously<sup>28</sup>. Genomic DNA was reconstituted into nucleosomes under selective pressure for nucleosome-favouring sequences by salt gradient dialysis<sup>13</sup>, using 40 µg histone octamer plus 100 µg DNA in a 200 µl volume. This results in a lower concentration of nucleosomes along the DNA than obtained *in vivo*. This technical limitation was necessitated by our finding that reconstitutions at the *in vivo* stoichiometry on this long genomic DNA resulted in insoluble chromatin that was inaccessible to micrococcal nuclease. Two independent reconstitutions were carried out. The resulting nucleosomes were biochemically isolated by MNase digestion, using  $6 \times 10^{-3}$  units MNase (Sigma Chemical Company) per 10 µg competitively reconstituted DNA, in 10 mM Tris, pH 8.0, and 1 mM CaCl<sub>2</sub> for 5 min at 37 °C. DNA was extracted, and protected fragments of length ~147 bp were isolated by polyacrylamide gel electrophoresis, and then extracted from the gel. Both samples were independently subjected to Illumina sequencing at the UNC high-throughput sequencing facility. We note that, because the chicken histones represent a mixture of post-translationally modified and unmodified histones as well as variant histones, there is a possibility that these modifications or variants influence the measured nucleosome positions.

**Mapping and post-processing of parallel sequencing reads.** We mapped reads resulting from sequencing experiments to the yeast genome, allowing at most one mismatch and no gaps. To estimate the mean DNA fragment length in each experiment, we superimposed the nucleosome reads of one strand and examined the distribution of nucleosome reads of the opposite strand. As expected, this distribution showed a strong peak at ~140–170 bp for all experiments, with slight variations between experiments. We used the maximum of the peak as an estimation of the mean DNA fragment length and extended all nucleosome reads to this length. We defined repetitive regions as regions that were matched by a read that mapped to more than one place in the genome. We excluded repetitive regions and their 150-bp vicinity from our analyses. To obtain genomic nucleosome occupancy tracks we summed for each position all reads covering it. We noticed that in some replicates from each of the three *in vivo* conditions (YPD, YPetOH and YPGal), but not in either of the *in vitro* replicates, a small fraction of the mapped base pairs (in all cases less than 0.5%) were covered by an aberrantly large number of reads (more than ten times the median genomic base pair coverage) that may potentially result from sequencing biases. For these regions, we set the coverage of these base pairs to equal ten times the median genomic base-pair coverage. Next, for each map, we calculated the average nucleosome occupancy at every base pair by dividing the number of reads that cover that base pair by the average number of reads per base pair across the genome. Finally, for each replicate we took the log of the ratio at every base pair and set the genomic mean to zero by subtracting the mean from every base pair. Thus, base pairs for which the occupancy value is below 0 showed a relative depletion of nucleosomes in comparison with the genome-wide average, whereas base pairs with values above 0 have a relative enrichment. The independent replicates for each experiment type were in excellent agreement, so we averaged the replicates within each type to create four nucleosome occupancy maps: three *in vivo* (YPD, YPetOH, YPGal) and one *in vitro*. These final tracks are termed the normalized nucleosome occupancy. The data in Fig. 1 are presented without the log transformation.

**Analysis of nucleosomes reconstituted on oligonucleotides.** Two concerns with our experimental approach were the known sequence specificity of micrococcal nuclease<sup>20</sup> and the possibilities of biases introduced by the parallel

DNA-sequencing strategy. We addressed both of these concerns simultaneously with an independent and different kind of experiment that obtains large-scale measurements of *in vitro* nucleosome sequence preferences without the use of micrococcal nuclease or parallel sequencing. In brief, we created a pool of ~40,000 double-stranded oligonucleotides of length 150 bp, each flanked by common priming sites<sup>29</sup>, and combined the pool with limiting amounts of chicken histones to form nucleosomes. DNAs in reconstituted nucleosomes were separated from unincorporated DNAs by native gel electrophoresis, without the use of any micrococcal nuclease. We then extracted from the gel the DNA that had successfully competed to form nucleosomes, and used both parallel sequencing and microarrays to compare the nucleosomal DNAs to DNAs in the initial pool. For each sequence, we calculated the log-ratio between the reconstituted fraction and the initial pool as a measure of the nucleosome affinity of that sequence. Thus, a log-ratio above zero indicates a higher than average affinity to form nucleosomes, whereas a value below zero indicates a lower than average affinity. We note that this experiment is inherently noisy, as shown by a large variation in the number of sequence reads in the unselected reference sample, which may be due to the use of two rounds of PCR in the procedure.

Specifically, we created 197-mer oligonucleotides (containing 150-mer unique sequences flanked by common priming sites) using a modification of the microarray-based method described previously<sup>29</sup>. Sequences were designed to represent tiled yeast genomic sequences (22,236 oligonucleotides), and a survey of randomly generated sequences selected to represent a wide variety of DNA sequence characteristics (nucleotide and dinucleotide composition, periodic features and DNA structural parameters<sup>7</sup>). DNA obtained from the array was PCR-amplified and assembled into nucleosomes under selective pressure for nucleosome-favouring sequences by salt gradient dialysis<sup>13</sup>, using 6 µg histone octamer and 12 µg DNA in a 200-µl volume. The resulting nucleosomes were separated away from the remaining naked DNA by native polyacrylamide gel electrophoresis (5% (w/v) polyacrylamide) in 1/3× TBE buffer (TBE is 90 mM Tris, 90 mM boric acid, 2 mM EDTA, pH 8.3). The nucleosome-containing band was excised from the gel, and its DNA extracted by crushing and soaking into 0.3 M NaOAc, 0.5 M NH<sub>4</sub>OAc, 0.1 mM EDTA and 0.1% (w/v) SDS, followed by ethanol precipitation. For Illumina analysis, the pools before and after nucleosome assembly were re-amplified and NruI sites present in the primers were cleaved before sequencing from the ends. In our analyses, we used oligonucleotides that were sequenced at least once and at most 500 times in each experiment. For microarray analyses, the pools before and after assembly were re-amplified with fluor-labelled primers, and hybridized to an Agilent array designed to detect the individual sequences represented in the pools. Additional details are available in the Supplementary Information.

To remove any direct biases resulting from the primers in the computation of preferences over sequences of length 5 from these data (Supplementary Fig. 7), we omitted the primers altogether in the computation of 5-base-pair preferences (that is, all 5-base-pair sequences were extracted only from the variable oligonucleotides without the primers). This left open the possibility of indirect biases in preferences of sequences of length 5 resulting from the primers; for example, if one primer sequence attracted nucleosomes and if there were systematic biases in compositions of sequences of length 5 near that end of the designed oligonucleotides. However, half (~20,000) of the oligonucleotides were selected from the yeast genome, and the other half were designed in several unrelated ways (see Supplementary Methods) and do not have such biases. The two approaches are in strong agreement regarding these preferences over sequences of length 5 base pairs, diminishing concerns regarding biases caused by the primers (the correlations between the preferences computed from the first approach and the second approach are 0.81 and 0.83 for reading out the measurements of the second approach by parallel sequencing or microarrays, respectively; see Supplementary Fig. 7).

**Data sets.** The yeast genome sequence (May 2006 build) and gene and chromosome annotations were obtained from SGD<sup>30</sup>. Yeast transcription start sites were compiled from refs 31–33: for each gene, the transcription start site was taken as that with the most sequence reads from refs 31 and 32, or from ref. 33 when no sequencing data was available. Functional transcription factor DNA binding sites in yeast, defined as sites that are bound by their cognate transcription factor were obtained from refs 22 and 34. Expression levels in yeast were taken from ref. 26. The *in vivo* map of nucleosome organization in *C. elegans* was taken from ref. 19; specifically, we used both the 'adjusted nucleosome coverage' and 'raw nucleosome coverage', applying the same normalization method to the latter as we did to our yeast nucleosome maps<sup>19</sup>.

**Classifying nucleosome-enriched regions from nucleosome-depleted regions.** To test whether the two maps agree on the scale of individual nucleosomes, we took the target map, and from it defined nucleosome-enriched regions as maximal consecutive regions longer than 50 bp for which minimum occupancy at every base pair is above some threshold,  $t_c$ , and nucleosome-depleted regions as



maximal consecutive regions longer than 50 bp for which maximum occupancy at every base pair is below some threshold,  $t_d$ . We chose 50 bp as a compromise between partitioning as much of the genome as possible into nucleosome-enriched and nucleosome-depleted regions while still deriving each partition from enough data to reduce noise that may arise when regions are too short (a choice of 25 bp yielded equivalent results). At various thresholds, we then scored each of the resulting nucleosome-enriched and nucleosome-depleted regions by its average occupancy in the predicting map. Note that at the threshold  $t_d = t_d = 0$ , the evaluation is done across the entire genome (typically ~97%, because regions <50 bp are excluded), and the median length of the nucleosome-enriched regions is typically 150–190 bp. If the predicting map were fully predictive of the target map, then the predicting map occupancy score of every 'enriched' region in the target map would be higher than that of every 'depleted' region in the target map. A standard quantification of this predictive power is the receiver operating characteristic curve, whose area under the curve is 1 for perfect performance and 0.5 for random guessing.

**Sequence-based model for nucleosome positioning.** We use the *in vitro* map to represent the sequence preferences of nucleosomes by devising a probabilistic model, similar in formulation to that proposed in ref. 17, that assigns a score to every 147-bp (nucleosome-length) sequence. Our model is based on both the position-dependent and global preferences of sequences of length 5 that we characterized previously (Fig. 3). The first component,  $P_N$ , represents the distribution over dinucleotides at each position along the nucleosome length, and thus captures the periodic signal of dinucleotides along the nucleosome. The second component,  $P_L$ , represents the position-independent distribution of nucleosomes over sequences of length 5, and thus captures sequences that are generally favoured or disfavoured by nucleosomes regardless of their detailed position within the nucleosome. We chose to represent this component using sequences of length 5, because this is the highest order  $k$ -mer for which our data have sufficient statistics to robustly estimate each of the associated parameters. The final score that our model assigns to a 147-bp sequence  $S$  is then given by the log-ratio of these two model components:

$$\begin{aligned} \text{Score}(S) &= \log \frac{P_N(S)}{P_L(S)} \\ &= \log \frac{P_{N,1}(S[1]) \prod_{i=2}^{147} P_{N,i}(S[i]|S[i-1])}{P_L(S[1]) \prod_{i=2}^{147} P_L(S[i]|S[\max(1, i-4)], \dots, S[i-1])} \end{aligned} \quad (1)$$

where  $P_{N,i}$  is the  $i$ -th component of the dinucleotide model component and represents the conditional probability distribution over nucleotides at position  $i$  given the nucleotide that appeared at position  $(i-1)$ , and  $P_L$  is the position-independent component of the second component of our model ( $P_L$ ). Note that  $P_{N,1}$  is represented by a mononucleotide model over the nucleotide at the first position.

We now describe in detail how each of the two components of our model was derived. To estimate the position-specific dinucleotide component of our model,  $P_N$ , we took all of the nucleosome reads (a total of over 10,000,000) from our collection of nucleosome-bound sequences, and first aligned these sequences about their centre. Each sequence was added twice to the alignment, once in its original form and once in its reverse complement form, to account for the two-fold symmetry in the nucleosome structure<sup>35</sup>. With each position  $i$ , we then associated a dinucleotide distribution,  $P_{N,i}$ , which we estimated from the combined dinucleotide counts at alignment positions  $[i-2, i-1]$ ,  $[i-1, i]$  and  $[i, i+1]$  (the two end positions of the nucleosome are averaged with fewer positions). Combining the dinucleotides at the two neighbouring positions smoothes the resulting dinucleotide distribution at each position with a 3-bp moving average, and is motivated by the experimental evidence that small  $\pm 1$ -bp changes in spacing of key nucleosome DNA sequence motifs can occur with a relatively small cost to the free energy of histone–DNA interactions<sup>36</sup>. To remove sequence composition biases from this component, we normalized the distribution over each position separately, by dividing the final probability of every dinucleotide at each position by the average probability of that dinucleotide across all positions, and finally normalized the resulting dinucleotide weights to a probability distribution. We used this estimation procedure in the 127 central positions of the nucleosome, and we forced a uniform distribution over the 10 remaining positions at each end of the nucleosome profile. This was done to avoid biases in nucleotide distributions that may arise from the sequence specificity of the micrococcal nuclease used to isolate the nucleosome, because this way we do not include statistics that are taken from the cut site of the

nuclease. Note that our above construction produces a reverse complement symmetrical distribution; that is, the probability of a sequence and its reverse complement are equal by construction.

The position-independent component of our model,  $P_L$ , the purpose of which is to represent sequences that are generally favoured or disfavoured regardless of their position within the nucleosome, assigns a score to each 147-bp sequence, as the product of a position-independent Markov model,  $P_L$ , of order 4. Thus,  $P_L$  defines a probability distribution over every one of the 1,024 possible sequences of length 5. We chose to model the distribution over sequences of length 5, because this is the highest order for which our data still provides sufficient statistics to robustly estimate the value of each of the 1,024 parameters. Given a map of nucleosome positions, the probability of each sequence of length 5 is equal to the reciprocal of the average nucleosome occupancy of that sequence across all instances of the 5-base-pair-long sequence in the map, where this reciprocal average occupancy is then scaled to a probability by dividing it by the sum of all such reciprocal occupancies across all 5-base-pair-long sequences (we use the reciprocal of the average occupancy so that higher probabilities correspond to disfavoured sequences). Thus, this Markov model,  $P_L$ , includes contributions both from sequences that are disfavoured by nucleosomes and from sequences that are favoured by nucleosomes, because it models the distribution over all sequences of length 5, with the disfavoured sequences having a relatively high probability and the favoured sequences having a relatively low probability.

The above probabilistic model assigns a nucleosome formation score to each sequence of (nucleosome-length) 147 bp. We then followed the approach of ref. 5, and used the scores of this model to compute the genome-wide distribution over nucleosome positions, taking into account steric hindrance constraints between neighbouring nucleosomes. These computations can be done efficiently using dynamic programming, and their output is a track of the average nucleosome occupancy per base pair, that is, the probability that each base pair in the genome is covered by any nucleosome.

In developing our model, we also separately evaluated the performance of the  $P_N$  and  $P_L$  components, and found that the performance of the  $P_L$  component alone is nearly the same as that of the full model (per-base-pair nucleosome occupancy correlation of 0.876 between a  $P_L$ -only model and the *in vitro* map, compared with 0.880 for the full model using both  $P_L$  and  $P_N$ ), whereas the performance of the  $P_N$  component alone is highly predictive but slightly worse than the full model (correlation of 0.820). Thus, for practical purposes of generating predictions, using only the  $P_L$  component may suffice. Nevertheless, the  $P_N$  component captures real periodic aspects of nucleosome sequence preferences that have been reported by several other groups<sup>4,6,37</sup>, and that we now find in the genome-wide *in vitro* nucleosome map in Fig. 3b. For these reasons, we retain the  $P_N$  component as part of the model.

For the synthetic oligonucleotide analysis (Supplementary Fig. 7), we used the nucleosome formation potential of equation (1). In all other analyses we converted this genome-wide nucleosome formation potential into genome-wide nucleosome occupancy per base pair, as described in ref. 5. We selected the concentration and temperature parameters ( $\tau$  and  $\beta$ , respectively) in a cross-validation manner to be  $\tau = 0.03$ ,  $\beta = 1$  in yeast and  $\tau = 0.1$ ,  $\beta = 1$  in *C. elegans*. Finally, to appropriately compare the model with log-transformed data, we applied to the predictions of the model the same log normalization that was applied to the data.

28. Feng, H. P., Scherl, D. S. & Widom, J. Lifetime of the histone octamer studied by continuous-flow quasielastic light scattering: test of a model for nucleosome transcription. *Biochemistry* **32**, 7824–7831 (1993).
29. Porreca, G. J. et al. Multiplex amplification of large sets of human exons. *Nature Methods* **4**, 931–936 (2007).
30. Cherry, J. M. et al. SGD: *Saccharomyces* genome database. *Nucleic Acids Res.* **26**, 73–79 (1998).
31. Zhang, Z. & Dietrich, F. S. Mapping of transcription start sites in *Saccharomyces cerevisiae* using 5' SAGE. *Nucleic Acids Res.* **33**, 2838–2851 (2005).
32. Miura, F. et al. A large-scale full-length cDNA analysis to explore the budding yeast transcriptome. *Proc. Natl Acad. Sci. USA* **103**, 17846–17851 (2006).
33. David, L. et al. A high-resolution map of transcription in the yeast genome. *Proc. Natl Acad. Sci. USA* **103**, 5320–5325 (2006).
34. Harbison, C. T. et al. Transcriptional regulatory code of a eukaryotic genome. *Nature* **431**, 99–104 (2004).
35. Richmond, T. J. & Davey, C. A. The structure of DNA in the nucleosome core. *Nature* **423**, 145–150 (2003).
36. Ong, M. S., Richmond, T. J. & Davey, C. A. DNA stretching and extreme kinking in the nucleosome core. *J. Mol. Biol.* **368**, 1067–1074 (2007).
37. Bolshoy, A., Ioshikhes, I. & Trifonov, E. N. Applicability of the multiple alignment algorithm for detection of weak patterns: periodically distributed DNA pattern as a study case. *Comput. Appl. Biosci.* **12**, 383–389 (1996).

# Transmembrane passage of hydrophobic compounds through a protein channel wall

Elizabeth M. Hearn<sup>1</sup>, Dimki R. Patel<sup>1</sup>, Bryan W. Lepore<sup>1</sup>, Mridhu Indic<sup>1</sup> & Bert van den Berg<sup>1</sup>

Membrane proteins that transport hydrophobic compounds have important roles in multi-drug resistance<sup>1–3</sup> and can cause a number of diseases<sup>4,5</sup>, underscoring the importance of protein-mediated transport of hydrophobic compounds. Hydrophobic compounds readily partition into regular membrane lipid bilayers<sup>6</sup>, and their transport through an aqueous protein channel is energetically unfavourable<sup>3</sup>. Alternative transport models involving acquisition from the lipid bilayer by lateral diffusion have been proposed for hydrophobic substrates<sup>3,4,7–12</sup>. So far, all transport proteins for which a lateral diffusion mechanism has been proposed function as efflux pumps. Here we present the first example of a lateral diffusion mechanism for the uptake of hydrophobic substrates by the *Escherichia coli* outer membrane long-chain fatty acid transporter FadL. A FadL mutant in which a lateral opening in the barrel wall is constricted, but which is otherwise structurally identical to wild-type FadL, does not transport substrates. A crystal structure of FadL from *Pseudomonas aeruginosa* shows that the opening in the wall of the  $\beta$ -barrel is conserved and delineates a long, hydrophobic tunnel that could mediate substrate passage from the extracellular environment, through the polar lipopolysaccharide layer and, by means of the lateral opening in the barrel wall, into the lipid bilayer from where the substrate can diffuse into the periplasm. Because FadL homologues are found in pathogenic and biodegrading bacteria, our results have implications for combating bacterial infections and bioremediating xenobiotics in the environment.

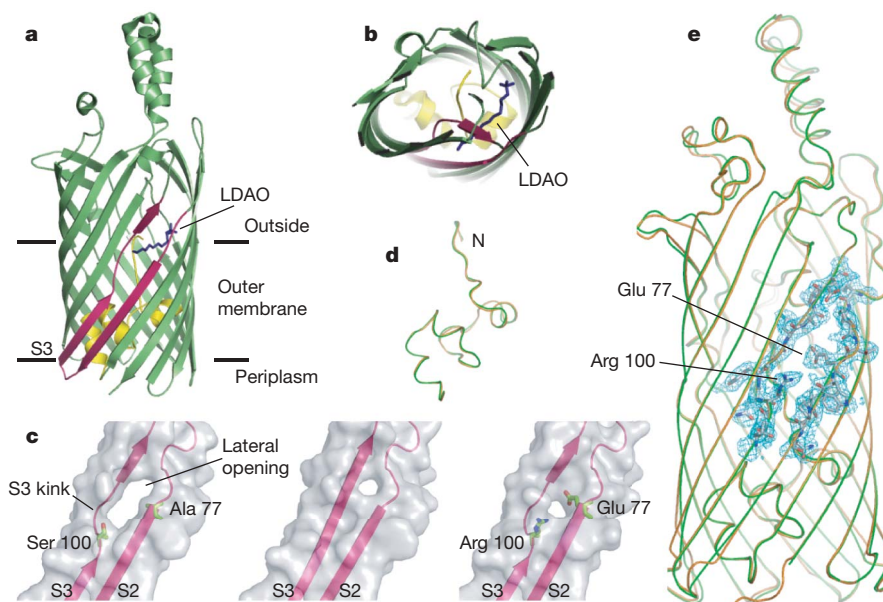
The outer membrane of Gram-negative bacteria provides an efficient barrier for the passage of hydrophobic molecules owing to the presence of the polar lipopolysaccharide (LPS) layer on the outside of the cell. The only protein family currently known to be involved in the uptake of hydrophobic molecules is the FadL family, named after the archetypal long-chain fatty acid (LCFA) transporter FadL of *E. coli*<sup>13,14</sup>. The crystal structures of *E. coli* FadL<sup>15</sup> revealed a monomeric 14-stranded  $\beta$ -barrel with an interior occluded by an amino-terminal hatch domain, and an inward-pointing kink in  $\beta$ -strand S3 that creates an unusual, lateral opening in the transmembrane barrel<sup>15</sup>. On the basis of the FadL structures, two possible LCFA transport mechanisms can be envisioned, both of which are diffusion-based because FadL-mediated transport does not require an energized membrane<sup>16</sup>. In the ‘classical’ transport model, the hatch undergoes conformational changes creating a transient channel for LCFA transport transverse to the membrane, from the extracellular medium directly to the aqueous periplasm<sup>15</sup>. The alternative ‘lateral diffusion’ transport model is based on the observation that an LCFA-mimicking LDAO detergent molecule is bound in the lateral opening of *E. coli* FadL (Fig. 1a, b). In the lateral diffusion model, LCFAs exit FadL laterally through the opening in the barrel wall to move into the outer membrane, from where they could diffuse into the periplasm. To answer the question regarding which

transport model is operative for FadL, we designed a number of FadL mutants focusing on the hatch domain and the lateral opening.

Two FadL mutants,  $\Delta$ S3 kink and A77E/S100R, were designed to close specifically the barrel wall opening in different ways. In  $\Delta$ S3 kink, four residues in strand S3 that line the lateral opening (<sup>100</sup>SNYG<sup>103</sup>) were replaced with three residues (AND) to bring strand S3 into register with the neighbouring strands and to destabilize the conformation of the S3 kink. In A77E/S100R, residues with long, oppositely charged side chains were introduced for residues Ala 77 and Ser 100, which are located in the barrel wall on different sides of the opening (Fig. 1c) with their side chains pointing towards each other. We anticipated that the introduced glutamic acid and arginine might form a salt bridge, closing the lateral opening. We determined the crystal structures of the  $\Delta$ S3 kink and A77E/S100R mutants (Supplementary Table 1). The  $\Delta$ S3 kink mutant, although largely identical to wild-type FadL, lacks density for several extracellular loops (Supplementary Fig. 1). However, the A77E/S100R structure is identical to that of wild-type FadL (Fig. 1d, e), with the sole exception of the lateral opening, which is much smaller in the A77E/S100R mutant (Fig. 1c). Thus, the introduced mutations for residues Ala 77 and Ser 100 are unlikely to affect the interior of the protein, strongly arguing against an inhibitory effect of the mutations on a possible conformational change in the hatch that would form a classical transport channel. Crucially, both the  $\Delta$ S3 kink and A77E/S100R mutants were completely inactive for *in vivo* LCFA uptake and growth on palmitate (Fig. 2). Moreover, the low transport activity (~3%) of the S100R single mutant (Fig. 2) demonstrates that the introduction of a single long side chain is sufficient to inhibit LCFA uptake efficiently. The structural and biochemical data together demonstrate that constricting the lateral opening in the FadL barrel wall is sufficient to block LCFA uptake. Although we can not exclude the possibility that subtle changes in barrel structure and dynamics, introduced by the mutations, affect the formation of a lateral opening in other parts of the barrel, our data are most easily explained by a transport model in which LCFAs diffuse laterally from the lumen of the barrel, through the observed opening in the barrel wall, into the outer membrane.

If transport by lateral diffusion is a general feature of FadL channels, then the opening in the barrel wall should be structurally conserved. To test this notion we have determined the crystal structure (Supplementary Table 2) of a FadL homologue from *Pseudomonas aeruginosa* (PaFadL), which has low (20%) sequence identity to *E. coli* FadL (EcFadL). Despite the modest sequence identity, PaFadL is structurally similar to EcFadL, with the exception of a number of extracellular loops (Fig. 3a, b). PaFadL has a pronounced lateral opening in the barrel wall at the same location as in EcFadL (Fig. 3a), indicating that the lateral opening is conserved in FadL family members. The most notable feature of the PaFadL structure is the presence of well-defined density for three complete tetraethylene glycol monooctyl

<sup>1</sup>Program in Molecular Medicine, University of Massachusetts Medical School, Worcester, Massachusetts 01605, USA.



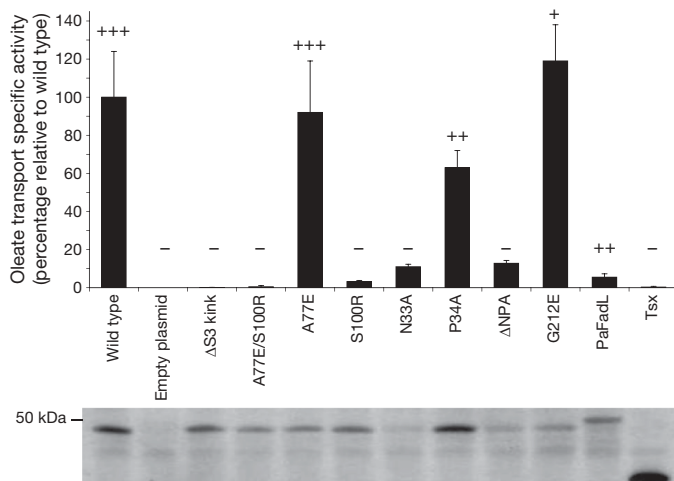
**Figure 1 | Structural features of FadL mutants.** **a, b,** Ribbon diagram of wild-type FadL highlighting the S3 kink, viewed laterally (**a**) and from the extracellular side (**b**). An LDAO molecule (dark blue) protrudes through the opening in the barrel wall between strands S2 and S3 (dark pink). The hatch domain is coloured yellow. The interface boundaries of the outer membrane bilayer are indicated by horizontal lines. **c,** Surface views of the kink region (with strands S2 and S3 indicated) of wild-type FadL (left),  $\Delta$ S3 kink

(middle) and A77E/S100R (right), showing the smaller lateral opening in  $\Delta$ S3 kink and A77E/S100R. The side chains for the residues at positions 77 and 100 are shown. **d, e,** Backbone superpositions of wild-type FadL (orange) and A77E/S100R (green; **d**, hatch domains only). In **e**,  $2F_o - F_c$  density (contoured at  $1.5\sigma$ ) is shown as a blue mesh for segments of  $\beta$ -strands S2 and S3 in A77E/S100R, with the side chains of residues Glu 77 and Arg 100 indicated.

ether ( $C_8E_4$ ) detergent molecules inside the barrel lumen (Fig. 3c). The first two detergent molecules (1, 2) are present at positions that are analogous to those occupied by detergent molecules in EcFadL<sup>15</sup>. The third  $C_8E_4$  molecule (3) is located in the lateral opening, analogous to the LDAO molecule in wild-type FadL (Fig. 1a, b). The presence of a detergent molecule in the lateral opening indicates that this part of the potential substrate passageway has a substantial affinity for LCFA

substrates. This is borne out by the fact that the environment of the detergent molecule in the opening is largely hydrophobic (Supplementary Fig. 2), like the other LCFA-binding sites in FadL<sup>15</sup>.

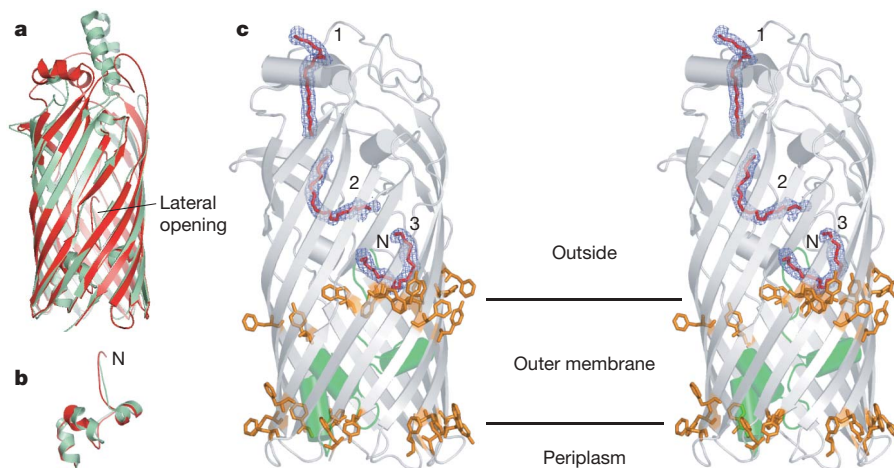
The three detergent molecules in PaFadL clearly delineate a long ( $\sim 55$  Å) passageway that runs from the extracellular surface all the way to the lateral opening in the barrel wall. Of the  $\sim 50$  residues that are located within 4.5 Å of the detergent molecules, more than 80% are hydrophobic (Supplementary Fig. 2). Moreover, structural and sequence alignments show that the hydrophobic character of these residues is conserved. The hydrophobic channel is continuous, and is interrupted only by the first three amino-terminal residues of the hatch, most notably Phe 3. Remarkably, a conformational change of the N terminus, as previously observed in EcFadL<sup>15</sup>, would generate an uninterrupted, hydrophobic passageway for substrate diffusion all the way from the extracellular medium to the lateral opening (Supplementary Fig. 3). As in EcFadL (Fig. 1a), the lateral opening in PaFadL is situated in the region of the polar–apolar interface of the outer leaflet of the outer membrane (Fig. 3c). This location makes sense, because it probably provides a favourable environment for both the carboxylate head group and the hydrocarbon chain as the LCFA emerges from the FadL lumen. Our observation is in agreement with recent data obtained for multidrug efflux pumps, where a range of hydrophobic substrates was shown to be preferentially localized in the interface region of the lipid bilayer<sup>6</sup>. Despite the structural similarity to EcFadL, PaFadL has very low activity for LCFA transport (Fig. 2), confirming that FadL channels are substrate specific<sup>17</sup>.



**Figure 2 | Functional analysis of FadL proteins.** Specific activity expressed relative to wild type for oleate uptake measured in whole cells at 20 °C. Bars represent the average of at least four independent measurements, and the error bars represent the standard deviation. Growth on palmitate minimal medium plates at 37 °C, ranging from wild-type levels (++) to no growth (-), is indicated above the bars. The hatch mutants N33A and  $\Delta$ NPA do not support growth of *E. coli* on palmitate at 37 °C, because the expression levels of these two mutants are below detection limits at this temperature. The bottom panel shows representative western immunoblots indicating the expression levels of the various proteins in the outer membrane at 20 °C. Tsx is the *E. coli* outer membrane nucleoside transporter, which is included as an additional negative control.

To support our lateral diffusion model, we also asked whether the hatch domain of EcFadL is flexible. The hatch NPA sequence (residues 33–35) is absolutely conserved and is a signature of FadL channels. The Asn 33 amide side chain of the NPA sequence forms a hydrogen bond with the Gly 21 backbone carbonyl, bringing distant parts of the hatch together (Supplementary Fig. 4). The removal of this hydrogen bond would seem a good candidate to probe whether the hatch domain can undergo conformational changes. We determined the X-ray crystal structures of the mutants N33A, P34A and  $\Delta$ NPA (N33G/P34G/A35G; Supplementary Table 1), and found that they are very similar to wild-type FadL (Supplementary Figs 1 and 4)





**Figure 3 | A hydrophobic passageway for substrate diffusion in PaFadL.** **a**, Superposition of EcFadL (green) and PaFadL (red), showing the conservation of the lateral opening. **b**, Superposition of the hatch domains. **c**, Stereo side view of PaFadL, with the three bound C<sub>8</sub>E<sub>4</sub> detergent molecules

indicated in red.  $2F_o - F_c$  density is shown as a blue mesh, contoured at  $2.0\sigma$ . The hatch domain is coloured green. The belts of aromatic residues that delineate the polar–apolar interfaces of the outer membrane are shown as orange stick models.

and do not have a hatch channel. The same is true for the G212E mutant (Supplementary Figs 1 and 4), which previously was proposed to have an open channel<sup>18</sup>. Consistent with the structural data, all hatch mutants are active in oleate transport (Fig. 2). The combined structural and biochemical data indicate that the NPA sequence is not directly involved in substrate transport, but it may be important for proper folding or outer membrane targeting of FadL. In addition, it appears that the hatch domain is rigid, providing support for the lateral diffusion transport model. It should also be noted that, although the hatch domain of EcFadL has a number of hydrophobic residues, these are interspersed with many polar residues (Supplementary Fig. 5). Therefore, even if a hatch channel in EcFadL could form by spontaneous conformational changes, it would not provide a suitably hydrophobic conduit for LCFA transport.

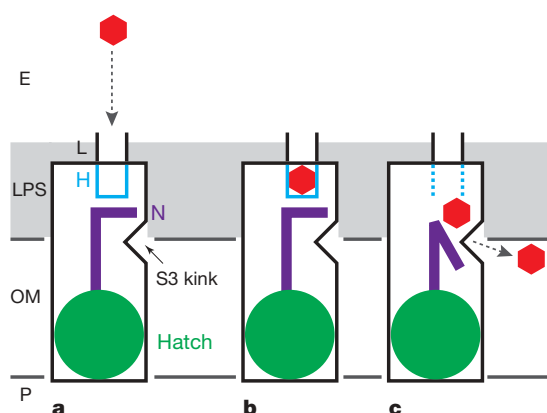
We now propose a general, lateral diffusion mechanism for the uptake of hydrophobic substrates by the FadL outer membrane protein family (Fig. 4). According to this mechanism, hydrophobic substrates diffuse by means of a long hydrophobic passageway laterally into the

outer leaflet of the outer membrane, probably by means of a stable, lateral opening in the barrel wall. From the outer leaflet the substrate can move to the inner leaflet of the outer membrane and diffuse into the periplasm. By using this mechanism, the hydrophobic substrates bypass the hydrophilic LPS layer of the outer membrane without having to move through an aqueous channel, which would be energetically unfavourable owing to the extremely low aqueous solubilities of FadL substrates ( $<0.1$  nM for palmitate<sup>19</sup>). Intriguingly, medium-chain fatty acids do not require FadL for uptake in *E. coli*<sup>20</sup>. Apparently these relatively water-soluble substrates ( $>0.5$  mM for laurate<sup>19</sup>) use other outer membrane channels for uptake, probably porins. Recently, crystal structures of *Pseudomonas putida* TodX and *Ralstonia pickettii* TbuX were determined, two members of a subfamily of FadL proteins from biodegrading bacteria involved in the uptake of mono-aromatic hydrocarbons<sup>21,22</sup>. As expected, TodX and TbuX both have a kink in strand S3 as well as a lateral opening at the same position as in FadL<sup>17</sup>, providing additional support for the generality of the lateral diffusion mechanism in FadL proteins. However, TodX and TbuX also have a narrow, continuous channel through the hatch domain that might serve as a classical channel. The TodX/TbuX hatch channel is relatively polar<sup>17</sup>, and may have evolved for the uptake of relatively water-soluble ( $\sim 1$ – $5$  mM)<sup>23</sup> mono-aromatic hydrocarbons in bacteria lacking porins. Future experiments will be required to establish whether the hatch channel in TodX/TbuX contributes to transport.

Structural features resembling the lateral opening of FadL channels have been observed in two other unrelated outer membrane proteins, PagP and OmpW. The lateral opening in the lipid A palmitoyl-transferase PagP may allow access of the lipid substrate to the active site of the enzyme<sup>24</sup>. OmpW belongs to a widespread family of eight-stranded  $\beta$ -barrels<sup>25</sup>, with members present in operons dedicated to the degradation of hydrophobic molecules such as naphthalene (NahQ)<sup>26</sup> and alkanes (AlkL)<sup>27</sup> in pseudomonads, indicating that they may form uptake channels for these compounds. Such uptake would have to occur by lateral diffusion, because the eight-stranded barrel lumen is too narrow to form a classical channel<sup>25</sup>. Thus, bacterial substrate uptake by lateral diffusion into the outer membrane may be a widespread phenomenon. Taking into account the known occurrence of lateral diffusion in multidrug transporters, this mode of transport probably represents a universal mechanism for membrane proteins that modify and transport hydrophobic substrates.

## METHODS SUMMARY

LCFA functional assays were performed using a modified approach from that described previously<sup>18</sup>. FadL mutant proteins were overexpressed and purified from *E. coli* C43 (DE3) as described for the wild-type protein<sup>15</sup>. Crystals were



**Figure 4 | Proposed lateral diffusion model for the uptake of hydrophobic substrates by FadL proteins.** **a**, Substrate (red hexagon) capture from the extracellular medium by a low-affinity binding site (L)<sup>15</sup>; **b**, diffusion of the substrate into an adjacent high-affinity binding site H (blue)<sup>15</sup>; **c**, spontaneous conformational changes in the N terminus (purple) result in substrate release and create a continuous passageway to the barrel wall opening formed by the kink in strand S3. The substrate diffuses laterally through the opening into the outer membrane (OM). The polar part of the LPS, constituting the principal barrier in the transport process, is shown in grey. The extracellular milieu (E) is at the top and the periplasm (P) is at the bottom.

obtained by the hanging-drop method, and structures were solved by molecular replacement using as a search model the monoclinic FadL structure (Protein Data Bank accession code 1T16) in which the hatch domain and the kink were deleted. PaFadL was overexpressed in native form in *E. coli* C43 (DE3), and purified as EcFadL. The PaFadL crystal structure was determined by the multiple isomorphous replacement with anomalous scattering (MIRAS) method, using a gold and osmium derivative obtained by the quick soaking method<sup>28</sup>.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 7 August; accepted 28 November 2008.**

**Published online 1 February 2009.**

- Saier, M. H. Jr & Paulsen, I. T. Phylogeny of multidrug transporters. *Semin. Cell Dev. Biol.* **12**, 205–213 (2001).
- Poole, K. Efflux pumps as antimicrobial resistance mechanisms. *Ann. Med.* **39**, 162–176 (2007).
- Sharom, F. J. Shedding light on drug transport: structure and function of the P-glycoprotein multidrug transporter (ABCB1). *Biochem. Cell Biol.* **84**, 979–992 (2006).
- van Meer, G., Halter, D., Sprong, H., Somerharju, P. & Egmond, M. R. ABC lipid transporters: extruders, flippases or flopless activators? *FEBS Lett.* **580**, 1171–1177 (2006).
- Borst, P., Zelcer, N. & van Helvoort, A. ABC transporters in lipid transport. *Biochim. Biophys. Acta* **1486**, 128–144 (2000).
- Siarheyeva, A., Lopez, J. J. & Glaubit, C. Localization of multidrug transporter substrates within model membranes. *Biochemistry* **45**, 6203–6211 (2006).
- Higgins, C. F. & Gottesman, M. M. Is the multidrug transporter a flippase? *Trends Biochem. Sci.* **17**, 18–21 (1992).
- Higgins, C. F. Multiple molecular mechanisms for multidrug resistance transporters. *Nature* **446**, 749–757 (2007).
- van Veen, H. W., Putman, M., Margolles, A., Sakamoto, K. & Konings, W. N. Structure–function analysis of multidrug transporters in *Lactococcus lactis*. *Biochim. Biophys. Acta* **1461**, 201–206 (1999).
- Bolhuis, H. *et al.* Multidrug resistance in *Lactococcus lactis*: evidence for ATP-dependent drug extrusion from the inner leaflet of the cytoplasmic membrane. *EMBO J.* **15**, 4239–4245 (1996).
- Shapiro, A. B. & Ling, V. Extraction of Hoechst 33342 from the cytoplasmic leaflet of the plasma membrane by P-glycoprotein. *Eur. J. Biochem.* **250**, 122–129 (1997).
- Shapiro, A. B. & Ling, V. Transport of LDS-751 from the cytoplasmic leaflet of the plasma membrane by the rhodamine-123-selective site of P-glycoprotein. *Eur. J. Biochem.* **254**, 181–188 (1998).
- Nunn, W. D. & Simons, R. W. Transport of long-chain fatty acids by *Escherichia coli*: mapping and characterization of mutants in the *fadL* gene. *Proc. Natl Acad. Sci. USA* **75**, 3377–3381 (1978).
- Black, P. N., Said, B., Ghosn, C. R., Beach, J. V. & Nunn, W. D. Purification and characterization of an outer membrane-bound protein involved in long-chain fatty acid transport in *Escherichia coli*. *J. Biol. Chem.* **262**, 1412–1419 (1987).
- van den Berg, B., Black, P. N., Clemons, W. M. Jr & Rapoport, T. M. Crystal structure of the long-chain fatty acid transporter FadL. *Science* **304**, 1506–1509 (2004).
- DiRusso, C. C. & Black, P. N. Bacterial long-chain fatty acid transport: gateway to a fatty acid-responsive signalling system. *J. Biol. Chem.* **279**, 49563–49566 (2004).
- Hearn, E. M., Patel, D. R. & van den Berg, B. Outer-membrane transport of aromatic hydrocarbons as a first step in biodegradation. *Proc. Natl Acad. Sci. USA* **105**, 8601–8606 (2008).
- Kumar, G. B. & Black, P. N. Linker mutagenesis of a bacterial fatty acid transport protein: identification of domains with functional importance. *J. Biol. Chem.* **266**, 1348–1353 (1991).
- Vorum, H., Brodersen, R., Kragh-Hansen, U. & Pedersen, A. O. Solubility of long-chain fatty acids in phosphate buffer at pH 7.4. *Biochim. Biophys. Acta* **1126**, 135–142 (1992).
- Black, P. N. Characterization of FadL-specific fatty acid binding in *Escherichia coli*. *Biochim. Biophys. Acta* **1046**, 97–105 (1990).
- Wang, Y. *et al.* Identification of a membrane protein and a truncated LysR-type regulator associated with the toluene degradation pathway in *Pseudomonas putida* F1. *Mol. Gen. Genet.* **246**, 570–579 (1995).
- Kahng, H. Y., Byrne, A. M., Olsen, R. H. & Kukor, J. J. Characterization and role of *tbuX* in utilization of toluene in *Ralstonia pickettii* PKO1. *J. Bacteriol.* **182**, 1232–1242 (2000).
- Eastcott, L., Shiu, W. Y. & Mackay, D. Environmentally relevant physical-chemical properties of hydrocarbons: a review of data and development of simple correlations. *Oil Chem. Pollut.* **4**, 191–216 (1988).
- Ahn, V. E. *et al.* A hydrocarbon ruler measures palmitate in the enzymatic acylation of endotoxin. *EMBO J.* **23**, 2931–2941 (2004).
- Hong, H., Patel, D. R., Tamm, L. K. & van den Berg, B. The outer membrane protein OmpW forms an eight-stranded beta-barrel with a hydrophobic channel. *J. Biol. Chem.* **281**, 7568–7577 (2006).
- Eaton, R. W. Organization and evolution of naphthalene catabolic pathways: sequence of the DNA encoding 2-hydroxychromene-2-carboxylate isomerase and *trans*-o-hydroxybenzylidenepyruvate hydratase-aldolase from the NAH7 plasmid. *J. Bacteriol.* **176**, 7757–7762 (1994).
- van Beilen, J. B., Eggink, G., Enequist, H., Bos, R. & Witholt, B. DNA sequence determination and functional characterization of the OCT-plasmid-encoded *alkJL* genes of *Pseudomonas oleovorans*. *Mol. Microbiol.* **6**, 3121–3136 (1992).
- Sun, P. D., Radaev, S. & Kattah, M. Generating isomorphous heavy-atom derivatives by a quick-soak method. Part I: test cases. *Acta Crystallogr. D* **58**, 1092–1098 (2002).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank the personnel of the National Synchrotron Light Source (NSLS) beamlines X6A and X29 for beam time and beamline support. We are grateful to P. Black and C. Petteys for the strain D10 and for their technical advice on the fatty acid transport assays. This work was supported by a training grant from the National Institutes of Health (to E.M.H.) and by a NIH research grant (1R01GM074824 to B.v.d.B.).

**Author Contributions** E.M.H. cloned, purified and crystallized FadL mutants, performed activity assays, and wrote the paper; D.R.P. cloned, purified and crystallized FadL mutants; B.W.L. performed activity assays; M.I. purified and crystallized PaFadL; and B.v.d.B. determined crystal structures, designed research and wrote the paper.

**Author Information** Coordinates and structure factors have been deposited in the Protein Data Bank under the following accession numbers: PaFadL, 3DWO;  $\Delta$ S3 kink, 2R88; A77E/S100R, 3DWN; P34A, 2R4L; N33A, 2R4N;  $\Delta$ NPA, 2R4O; and G212E, 2R4P. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to B.v.d.B. ([bert.vandenberg@umassmed.edu](mailto:bert.vandenberg@umassmed.edu)).

## METHODS

**FadL mutagenesis.** The *E. coli fadL* gene, including the signal sequence and a carboxy-terminal hexahistidine tag, was cloned as previously described<sup>15</sup> into the pBAD22 vector, which is under the control of the arabinose-inducible promoter<sup>29</sup>. Mutations were introduced into the *fadL* gene by using the QuikChange Site-directed Mutagenesis kit (Stratagene), and the mutations were verified by nucleotide sequencing.

**LCFA functional analyses.** For the functional analyses, *E. coli* LS6164  $\Delta fadR \Delta fadL$ <sup>30</sup> was transformed with the pBAD22 plasmids carrying the mutant *fadL* genes and FadL homologues. As a negative control, the pBAD22 plasmid carrying the outer membrane nucleoside transporter Tsx<sup>31</sup> was introduced into *E. coli* LS6164. The ability of the mutant proteins to support growth on LCFAs was measured by plating cells ( $5 \times 10^6$  colony-forming units per ml) on agar plates containing 5 mM sodium palmitate (Sigma), 0.5% (w/v) Brij 58, M9 minimal medium, and 1.5% (w/v) Noble agar (Difco). Growth was scored after 96-h incubation at 37 °C.

Transport assays with radiolabelled oleic acid were performed with modifications to the method described previously<sup>18</sup>. *E. coli* LS6164  $\Delta fadR \Delta fadL$  cells with the FadL mutant and pBAD22 plasmids were grown in Luria Bertani medium to mid-log phase, and FadL expression was induced with 0.01% (w/v) arabinose at 20 °C for 5 h. Cells were collected, washed in EB1 buffer (10 mM citric acid, 0.8 mM magnesium sulphate, 20 mM sodium ammonium phosphate, 60 mM potassium phosphate, 0.02 mM thiamine), and resuspended to a  $D_{600}$  of 1 in EB1 buffer with 0.5% (w/v) Brij 58. After 30 min starvation at 37 °C, samples were taken for western immunoblotting analysis and also for measurement of oleic acid transport activity. The transport assay was performed by diluting the cells in EB1 buffer containing 20 mM glucose, 0.5% (w/v) Brij 58, and 0.02  $\mu$ Ci [<sup>3</sup>H-9,10]-oleic acid (Sigma, specific activity 30.0–50.5  $\mu$ Ci mmol<sup>-1</sup>). At 25 min, samples were removed and filtered through 0.45- $\mu$ m membrane filters (Metricel GN-6, Pall Life Sciences). Filters were washed with EB1 buffer containing 0.5% Brij 58, and the radioactivity retained on the filters was counted. Counts for the cells expressing the FadL protein were corrected for the background radioactivity associated with cells containing an empty pBAD22 plasmid. To determine the amount of FadL protein expressed in the outer membrane, cells were incubated with BugBuster (Pierce) while shaking for 30 min at 37 °C, and centrifuged for 10 min at 20,000g. Proteins were detected by western immunoblotting using the Penta-His-horseradish peroxidase conjugate antibody (Qiagen) and the ECL chemiluminescence detection kit (GE Healthcare). Proteins were quantified against FadL standards using the ImageJ program<sup>32</sup>. The transport activities of the cells expressing mutant FadL proteins were corrected for FadL expression levels, and the rates were expressed as a percentage of wild type.

**Purification, crystallization and structure determination of FadL mutants.** For crystallization, the FadL mutant proteins were expressed from the pBAD22 plasmids in *E. coli* C43 (DE3) cells<sup>33</sup> grown in 2 $\times$  YT medium by induction with 0.2% (w/v) arabinose for 6 h at 30 °C. Cells were collected, and the proteins were purified from the total membrane fraction as previously described for wild-type FadL<sup>15</sup> by Ni affinity chromatography in LDAO, gel filtration chromatography in LDAO (Anatrace), and a final gel filtration chromatography step in C<sub>8</sub>E<sub>4</sub> (Sigma). The purified proteins were concentrated to 5–10 mg ml<sup>-1</sup> and flash-frozen in liquid nitrogen.

Crystals were obtained by the hanging-drop method at 22 °C using commercially available screens (The Classics and MB Class II, Qiagen) or in-house screens. The mutant FadL proteins crystallized under the following conditions: N33A, 0.1 M cadmium chloride, 0.1 M sodium acetate (pH 4.6), 30% (v/v) PEG 400; P34A, 0.2 M ammonium acetate, 0.1 M sodium acetate (pH 4.6), 30% (w/v) PEG 4000;  $\Delta$ NPA, 0.05 M magnesium acetate, 0.05 M cacodylate (pH 5.5), 35% (w/v) PEG 2000; G212E, 0.2 M zinc acetate, 0.1 M cacodylate (pH 6.5), 18% (w/v) PEG 8000;  $\Delta$ S3 kink, 0.2 M ammonium sulphate, 0.1 M MES (pH 6.5), 30% (w/v) PEG 5000 MME; and A77E/S100R, 0.1 M sodium chloride, 0.1 M citrate (pH 5.6), 16% PEG 4000. The crystals were flash frozen (100 K) in the reservoir solution containing C<sub>8</sub>E<sub>4</sub> and 15–25% (v/v) glycerol by plunging in liquid nitrogen.

Diffraction data were obtained on beamline X6A at the National Synchrotron Light Source (Brookhaven National Laboratory). Data sets were integrated and scaled using HKL2000 (ref. 34). The structures of the mutant proteins were solved by molecular replacement using Phaser<sup>35</sup>; the monoclinic FadL structure

(PDB accession number 1T16) without the N-terminal 40 amino acids and the S3 kink residues 99–108 was used as the search model. Model building was done using COOT<sup>36</sup>, and refinement was done with CNS<sup>37</sup>. Data collection and refinement statistics for the mutant FadL proteins are summarized in Supplementary Table 1.

**Cloning of PaFadL.** The signal sequence cleavage site for PaFadL was predicted to occur between residues 20 and 21 using the SignalP program<sup>38</sup>. For expression and outer membrane localization of PaFadL, the mature gene (lacking the endogenous signal sequences and the first two residues, and additionally modified with a C-terminal hexahistidine tag) was amplified by PCR from *P. aeruginosa* PAO1 (ATCC 47085) genomic DNA. The mature gene fragment was cloned in-frame with the *E. coli* FadL signal sequence (plus the first two residues of the mature FadL sequence) into the pBAD22 vector<sup>29</sup>.

**Purification, crystallization and structure determination of PaFadL.** PaFadL was expressed in *E. coli* C43 (DE3) cells<sup>33</sup> and purified using the detergents LDAO and C<sub>8</sub>E<sub>4</sub> as described for *E. coli* FadL<sup>15</sup>. Crystallization trials of PaFadL were set up using the hanging-drop technique with commercially available crystallization screens (The Classics and MB Class II screens, Qiagen). Crystals for the PaFadL were obtained at 22 °C in 0.1 M lithium sulphate, 0.1 M MES buffer (pH 6.5), and 28–32% (w/v) PEG 400. The crystals typically appeared after 10 days and grew to their full size ( $\sim 15 \times 30 \times 100 \mu$ m) in 3–4 weeks. They belonged to space group C222<sub>1</sub>, diffracted to a resolution of 2.1 Å, and contained one molecule in the asymmetric unit (Matthews coefficient  $V_M \sim 4.2 \text{ Å}^3 \text{ Da}^{-1}$ , corresponding to  $\sim 55\%$  (v/v) solvent content<sup>39</sup>). The crystals were flash frozen (100 K) directly from the drop by plunging in liquid nitrogen.

A native data set was collected, as well as data sets for crystals that were soaked for  $\sim 40$  min in 10 mM OsCl<sub>3</sub> or KAuCl<sub>4</sub>. The data sets were processed using HKL2000 (ref. 34), showing that the osmium and gold crystals were isomorphous with the native crystals (Supplementary Table 2). Two osmium and two gold sites were found by multiple isomorphous replacement in SOLVE (Z-score 10.2; FOM 0.39 for data between 20 Å and 2.5 Å)<sup>40</sup>. Five additional gold and three osmium sites were found in isomorphous/anomalous difference maps and refined using SHARP<sup>41</sup>. Electron density maps obtained from SHARP were used for automatic model building in RESOLVE<sup>40</sup>, which resulted in  $\sim 60\%$  of the model being built. Further model building was done manually in COOT<sup>36</sup>, followed by refinement in CNS 1.2 (ref. 37). Data collection and refinement statistics for PaFadL are summarized in Supplementary Table 2.

29. Guzman, L. M., Belin, D., Carson, M. J. & Beckwith, J. Tight regulation, modulation and high-level expression by vectors containing the arabinose P<sub>BAD</sub> promoter. *J. Bacteriol.* **177**, 4121–4130 (1995).
30. Ginsburgh, C. L., Black, P. N. & Nunn, W. D. Transport of long-chain fatty acids in *Escherichia coli*: identification of a membrane protein associated with the *fadL* gene. *J. Biol. Chem.* **259**, 8437–8443 (1984).
31. Ye, J. & van den Berg, B. Crystal structure of the bacterial nucleoside transporter Tsx. *EMBO J.* **23**, 3187–3195 (2004).
32. Abramoff, M. D., Magelhaes, P. J. & Ram, S. J. Image processing with ImageJ. *Biophotonics Int.* **11**, 36–42 (2004).
33. Miroux, B. & Walker, J. E. Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J. Mol. Biol.* **260**, 289–298 (1996).
34. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
35. McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C. & Read, R. J. Likelihood-enhanced fast translation functions. *Acta Crystallogr. D* **61**, 458–464 (2005).
36. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
37. Brunger, A. T. et al. Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
38. Bendtsen, J. D., Nielsen, H., von Heijne, G. & Brunak, S. Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* **340**, 783–795 (2004).
39. Matthews, B. W. Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491–497 (1968).
40. Terwilliger, T. C. & Berendzen, J. Automated MAD and MIR structure solution. *Acta Crystallogr. D* **55**, 849–861 (1999).
41. de La Fortelle, E. & Bricogne, G. Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multi-wavelength anomalous diffraction methods. *Methods Enzymol.* **276**, 472–494 (1997).



# naturejobs

**THE CAREERS  
MAGAZINE FOR  
SCIENTISTS**

**W**hen President Barack Obama lifted the ban on US federal funding of embryonic stem-cell research on 9 March, he did more than make a symbolic gesture in favour of advancing science; and he did more than take a step towards long-touted (if not guaranteed) disease cures. His statement had immediate economic repercussions. In the near term, there is little chance that freeing up the research will bring cures — but there is reason to believe the move will bring near-term jobs and investment.

Obama said the administration of George W. Bush had forced a “false choice between sound science and moral values”. Science and morality are not inconsistent in this case, he argued, saying that embryonic stem-cell research is part of a humane effort to relieve human suffering. The president went on to imply that lifting the ban is part of a larger plan to give scientists the support they deserve. Scientists should be shielded from unnecessary political influence so that they can do their jobs “free from manipulation or coercion”.

The decision, however, is also about winning the war for top scientific talent. California’s efforts to do so are well known: its US\$3-billion stem-cell initiative, although struggling of late because of the state’s budget woes, has attracted some accomplished biologists. Other states have tried to follow suit, although they may now scale back their own funding programmes to make way for government funding (see *Nature* **458**, 130–131; 2009). Countries from South Korea to the United Kingdom to Australia have pursued not only the research results, but the researchers as well.

Obama did note the upshot of scant government support. “Promising avenues go unexplored,” he said. “Some of our best scientists leave for other countries that will sponsor their work. And those countries may surge ahead of ours in the advances that transform our lives.” Obama’s announcement intensifies the race for cures and reaffirms support for a field. It also raises the stakes in the race to find the best stem-cell biologists on the planet.

**Gene Russo is editor of *Naturejobs*.**

## CONTACTS

**Editor:** Gene Russo

**Assistant editor:** Karen Kaplan  
e-mail: [naturejobseditor@nature.com](mailto:naturejobseditor@nature.com)

**European Head Office, London**  
The Macmillan Building,  
4 Crinan Street, London N1 9XW, UK  
Tel: +44 (0) 20 7843 4961  
Fax: +44 (0) 20 7843 4996  
e-mail: [naturejobs@nature.com](mailto:naturejobs@nature.com)

**European Sales Manager:**  
Dan Churchward (4966)  
e-mail: [d.churchward@nature.com](mailto:d.churchward@nature.com)  
**Assistant European Manager:**  
Nils Moeller (4953)

**Natureevents:**  
Ghislaine Ababou (+44 (0) 20 7014 4015)  
e-mail: [g.ababou@nature.com](mailto:g.ababou@nature.com)

**Southwest UK/RoW:**  
Alexander Ranken (4944)

## Northeast UK/Ireland:

Matthew Ward (+44 (0) 20 7014 4059)

## France/Switzerland/Belgium:

Muriel Lestringuez (4994)

## Scandinavia/Spain/Portugal/Italy:

Evelina Rubio-Hakansson (4973)

## North Germany/The Netherlands/Eastern

Europe: Kerstin Vincze (4970)

## South Germany/Austria:

Hildi Rowland (+44 (0) 20 7014 4084)

## Advertising Production Manager:

Stephen Russell

To send materials use London address above.

Tel: +44 (0) 20 7843 4816

Fax: +44 (0) 20 7843 4996

e-mail: [naturejobs@nature.com](mailto:naturejobs@nature.com)

**Naturejobs web development:** Tom Hancock

**Naturejobs online production:** Dennis Chu

## US Head Office, New York

75 Varick Street, 9th Floor,  
New York, NY 10013-1917

Tel: +1 800 989 7718

Fax: +1 800 989 7103

e-mail: [naturejobs@nature.com](mailto:naturejobs@nature.com)

**US Sales Manager:** Ken Finnegan

## India

Vikas Chawla (+91 1242881057)

e-mail: [v.chawla@nature.com](mailto:v.chawla@nature.com)

## Japan Head Office, Tokyo

Chiyoda Building, 2-37 Ichigayatamachi,  
Shinjuku-ku, Tokyo 162-0843

Tel: +81 3 3267 8751

Fax: +81 3 3267 8746

## Asia-Pacific Sales Manager:

Ayako Watanabe (+81 3 3267 8765)

e-mail: [a.watanabe@natureasia.com](mailto:a.watanabe@natureasia.com)

## Business Development Manager, Greater

China/Singapore:

Gloria To (+852 2811 7191)

e-mail: [g.to@natureasia.com](mailto:g.to@natureasia.com)

# Personnel dilemma



The hiring of US science faculty members has slowed considerably.

**Karen Kaplan** tracks the administrators who are trying to keep programmes intact.

**T**he stain of the international economic downturn has spread worldwide, but science institutions in the United States seem to be feeling its effects more severely than those in other nations. And it is not clear whether the science portion of the \$787-billion federal economic stimulus package — \$10.4 billion of which is slated for the National Institutes of Health and \$3 billion for the National Science Foundation — will help schools mitigate their recruitment woes (see *Nature* doi:10.1038/news.2009.126; 2009).

For now, recruitment efforts are rapidly decelerating virtually across the board, as institutions struggle to expand — or, in some cases, simply maintain — their science faculties (see 'Business as usual' and 'House-bound'). It has prompted some university chancellors and provosts to nervously wonder just how long the recession's effects will last and whether their universities' core mission may start to decay.

State universities have absorbed the hardest blow. Stony Brook University in New York had a 10% across-the-board budget cut. "As a result, we've had to restrict hiring dramatically," says provost Eric Kaler. "In a good year, we might recruit and hire a dozen tenure-track faculty in the sciences. This year, we will probably do one." Consequences could range from increased teaching loads for faculty members to fewer lab sections and specialized graduate courses. "It means," Kaler says, "that maintaining our level of excellence is more difficult."

The effects have been felt nationwide. Stanford University in Palo Alto, California, has cancelled or delayed recruitment at its various schools, a university spokeswoman says. Similarly, Johns Hopkins University in Baltimore, Maryland, in February imposed a university-wide hiring freeze (except at the university's Applied Physics Laboratory), effective until June 2010. Clemson University in South Carolina also imposed a university-wide hiring freeze in October last year.



University of Virginia  
vice-president Tom Skalak.

Clemson University's cuts have already affected the university's science programmes, including reductions in the number of research animals for its agricultural programme. Teaching duties could be another casualty. Mark Schlissel, dean of biological sciences at the University of California, Berkeley, says he hopes to keep the teaching programme intact while waiting for the economy to improve. The university has significantly curtailed its recruitment efforts.

## Investing in people

Where endowments are more robust, and hence interest earned on investments still healthy, outlooks are brighter. Tom Skalak, vice-president for research at the University of Virginia in Charlottesville, is sanguine about a faculty recruitment slowdown of 10–20% this year. The university has other sources of revenue besides the state, including philanthropic and corporate contributions, as well as an endowment that has fared relatively well. "There's been a slowdown but we're able to hire," Skalak says.

"We're hiring ten new science faculty this year, and that's firm," says David Dolling, dean of the

IMAGES.COM/CORBIS

T. COGILL



## BUSINESS AS USUAL

In the United States, university recruitment is limping along, largely as a result of recession-induced shrinking endowments and dwindling state coffers. But some other countries' science programmes are not yet suffering the same fate. "So far, it's business as usual," says a spokesperson for the University of Oxford, UK.

The University of Exeter, also in Britain, is recruiting 29 science faculty members, according to Janice Kay, a deputy vice-chancellor. Targeted disciplines include climate change, extrasolar planets, functional materials, systems biology and translational medicine. "We are recognizing a very severe recession," she says, "but we're not going to be diverted from our mission of expanding first-class research at Exeter."

Germany's University of Kiel is insulated by virtue of



**University of Exeter deputy vice-chancellor Janice Kay (left) and University of Toronto vice-provost Edith Hillan.**

an agreement with the state of Schleswig-Holstein that guarantees it will receive five years of funding, some of which is devoted to recruitment, according to Susanne Schuck-Zöller, head of the university's communications. And the downturn has not affected scientist recruitment at the country's Max Planck institutes. "Our budget has not been pruned," says Felicitas von Aretin, head of science communication.



Nor has science-faculty recruitment faltered at the University of Toronto in Canada, although the university isn't hiring recklessly. "Divisions are being a bit more cautious," says Edith Hillan, vice-provost for academic affairs, noting that tuition revenues and federal funding continue to be strong. "If there's no absolute need to hire, then they're making the decision to wait another year." **K.K.**

school of engineering and applied science at George Washington University in Washington DC. Dolling says a conservative investment strategy has protected the university's endowment from complete collapse. "Our institutional prerogative is that we are growing engineering and sciences."

Peter Lepage, dean of the college of arts and sciences at Cornell University in Ithaca, New York, says his university is relatively unscathed, at least for now, thanks largely to an aggressive recruiting and hiring strategy in recent years and other funding sources besides state monies. "The size of my faculty has grown by 6% or 7%," he says. "So even if we rolled it back, we would be setting ourselves back a decade, not to the nineteenth century."

Like several other institutions that are expanding their faculties, Northeastern University in Boston, Massachusetts, is working to leverage its new hires by making sure they can operate across disciplines. "We



**Cornell University dean Peter Lepage.**

some candidates, for example, have declined offers because they can't sell their homes. "Faculty at all levels, from junior to senior, are finding it very, very difficult to accept our offers because they cannot divest themselves from their current living arrangements," Easterling says. "They can't sell their houses — especially if they're

from an area hard hit by the sub-prime mortgage market."

The downturn threatens to exacerbate the notorious 'two-body' problem — handling a partner's job needs as well as the recruit's. Spouses of recruits are baulking at leaving their current position if the recruiting university can't offer them a job. Recently,

Easterling says, Pennsylvania State University extended an offer to a petroleum engineer at a Canadian university. The engineer seemed receptive. But he soon revealed that he could not sell his home, and that his wife, a pharmacist, was concerned about finding a new job in State College. "He turned us down," says Easterling. **K.K.**

might have biologists working with sociologists," says provost Stephen Director, referring to the university's multidisciplinary 'strategic initiative', which helps spread out costs as well as strengthen the university's teaching programme. "We're trying," Director says, "to build our core strengths." Northeastern University has 46 faculty positions to fill, about half of which are in basic science.

## Adapt or die

At Pennsylvania State University in State College, hiring for physical-science faculty members has an emphasis on inter-disciplinarity, according to William Easterling, dean of the college of Earth and mineral sciences. "As long as we're willing to hire physical scientists who are adept at working across disciplines, then we're able to continue with our hiring," he says.

Some universities actually see the slowdown as an opportunity to collect top talent. Despite a serious endowment dip, Harvard University in Cambridge, Massachusetts, is seeking more than 50 new faculty, including science posts, said president Drew Faust in an announcement to alumni. Faust said that recruitment has slowed after a number of years of faculty growth, but that the university will continue to "plan for intensified efforts in select areas of academic priority".

Although Tufts University in Boston, Massachusetts, has put some faculty searches on hold, it is actively recruiting 52 faculty members, some of whom are in the sciences, according to spokeswoman Kim Thurler. "We see opportunities to attract outstanding faculty," she says.

A prolonged downturn could diminish any sense of optimism, however. "One year is not enough to hurt," says Steve Halperin, dean of the college of computer, mathematical and physical sciences at the University of Maryland in College Park. "But over a period of multiple years, if there's a significant [faculty] decrease, then I'm looking at a future with fewer faculty, and that will change a lot of how we do business." Halperin is expecting to hire two or three faculty members this year, down from a high of seven or eight during a good year.

Some predict even more dire consequences. An extended downturn could affect "new thinking, new ideas, new life in the nation's graduate programmes", says Robert Sowell, vice-president for programmes and operations at the Council of Graduate Schools in Washington DC. "The United States' position relative to the rest of the world could start slipping," he says. "I don't think this is a situation we can't recover from, but the real concern is our global competitiveness."

**Karen Kaplan is assistant editor of Naturejobs.**

## HOUSE-BOUND

Some US universities report that recruitment has been hampered by more than budget woes. The wider implications of the recession have hit potential recruits and affected their ability to accept positions. William Easterling, dean of the college of Earth and mineral sciences at Pennsylvania State University in State College, says



# MOVERS

**Louise Ryan, chief, mathematical and information sciences, Commonwealth Scientific and Industrial Research Organisation, Sydney, Australia**



**2006–09:** Chair, Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts

**1985–2009:** Assistant to full professor, Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts

Louise Ryan's humanitarian spirit resides in the mind of a mathematician. After recognizing her affinity for solving real-world problems, Ryan carved a path as a biostatistician assessing environmental risks to human health. As she returns to her native Australia to head the maths and information-sciences division for the Commonwealth Scientific and Industrial Research Organisation, she plans to create opportunities for like-minded students.

As an undergraduate, Ryan began actuarial studies at Macquarie University in Sydney, Australia. She lost interest in the financial focus at the same time that she became captivated by statistics. "The real world is chaotic, and I loved coming up with simple formulae able to reveal the complexity of randomness," she says.

Ryan's undergraduate adviser, Don McNeil, encouraged her to pursue a PhD in the United States. At Harvard University, Ryan developed diagnostic techniques to determine which statistical models are most appropriate for specific data sets, but this focus was too theoretical to sustain her interest. Through McNeil, she met members of Harvard's biostatistics department and began a postdoc with Stephen Lagakos, who was working to quantify cancer risks of red dye 40, a food-colouring agent. At the same time, she worked on statistical underpinnings of clinical-trial design at the nearby Dana-Farber Cancer Institute.

Ryan's reputation for conducting quality environmental risk-assessments got her a seat on the National Academy of Sciences (NAS) panel on arsenic in drinking water, where she used her statistical expertise to quantify the effects associated with models' uncertainty. She used similar techniques to help a NAS panel advise on methylmercury. "Louise cares deeply about making the world a better place," says Lagakos.

While at Harvard, Ryan headed student programmes, including one that gave minority students opportunities in biostatistics. After directing an environmental statistics programme, she went on to chair the biostatistics department. She remains an adjunct professor at Harvard.

Her main goal now is stimulating top-notch statistics research opportunities in Australia. Already, she has started a graduate fellows programme to attract talented undergraduates who want work experience on the way to a PhD — a missing link in Australia's maths career pipeline. "I want to make this division a magnet for the best talent," Ryan says.

Virginia Gewin

## NETWORKS & SUPPORT

### A bill to care for carers

Many scientists, especially women, are familiar with the disruption to their career and funding that can occur as a result of extended family leave. To address this and the discrimination that sometimes results, Congresswoman Eddie Bernice Johnson (Democrat, Texas) has introduced a bill seeking grant extensions for federally funded scientists who take time off to care for family members. Funding agencies would also provide money to hire temporary workers.

Johnson based the bill on recommendations in the 2007 National Academies report *Beyond Bias and Barriers*. This is her second attempt; last year's bill died in committee. With the new US administration, she says, "I think we might be trying at a better time now".

The bill calls for workshops to educate university leaders, grant reviewers and other federally funded researchers about how to reduce gender bias. Some universities have balked at the suggested workshops' time-consuming data-collection requirements, which include monitoring the careers of female faculty members from workshop participants' departments.

Many institutions have not compiled such data before, says Joan

Herbers, a biologist at Ohio State University in Columbus on a National Science Foundation grant aimed at boosting participation of women in science and engineering.

Some agencies, including the National Institutes of Health, already extend grants while investigators are caring for family. Provisions for temporary workers, however, would be new, says Mary Ann Mason, co-director of the Center on Health, Economic and Family Security at the University of California, Berkeley. Although researchers on multi-year grants cannot be easily replaced, Mason says hiring a temporary worker may help in some cases, and could limit concerns about hiring women of child-bearing age. "It's not an easy fix by any means," she says. "But it's a necessary policy."

The bill's provisions are not limited to women. Men are increasingly requesting paternity leave, notes Phoebe Leboy, president of the Association for Women in Science. "The best thing that can happen for women in science," she says, "is for men in science to have similar responsibilities and a better understanding of what the problems are." Johnson hopes for action this congressional session.

Heidi Ledford

#### POSTDOC JOURNAL

### As simple as possible

Sitting in an empty lab, surrounded by a dark campus, I think to myself: "Why do I want to be a scientist?" I recall a story my father told me. Unclear about the practical application of a difficult measurement, someone asked physiologist A. V. Hill: why did you do the experiment? Hill paused, then said: "Because it amused me." I want to be a scientist to satisfy my curiosity about the world.

Staring at my first assistant-professor application, I think of what amuses me and satisfies my curiosity. I started my current project with a complex conceptual model for muscle activation. Recalling Henry David Thoreau's advice to "simplify, simplify", I stripped away the complexity with careful assumptions. Then I found an analytical expression whose taut curve hugged the numerically generated points of the complex model. This process of making a problem, in Einstein's words, "as simple as possible, but not simpler" is what amuses me.

I send my application into cyberspace and wonder about my prospects. Late last year, theoretical and applied mechanics at Cornell, the department where I received my PhD, ceased to exist. The dean merged it into the more experiment-driven (and more fundable) mechanical engineering department. Is there a place for someone whose passion lies more on the theoretical side? Must I ignore what amuses me in favour of what makes me marketable?

Sam Walcott is a postdoc in theoretical biophysics at Johns Hopkins University in Baltimore, Maryland.

# Takeaway

A word to the wise.

**Tony Ballantyne**

Stephanie stretched the phone cord to its full extent and leant around the doorway, gazing into the tiny kitchen beyond.

"There's a customer on the phone wanting two sweet-and-sour porks delivered."

Mr Ho shook fried rice from his wok, expertly filling a silver container.

"Why you telling me that, Stephanie? Can't you see I'm busy?"

"I know, Mr Ho, but the customer wants the order delivered to the anthill on Stonebreak Lane."

Mr Ho frowned. "Ah, that is unusual. Ants usually prefer lemon chicken. Tell them ten minutes."

Stephanie looked at Mr Ho. It wasn't like him to make jokes. It wasn't like him to do anything except to cook and to chastise her for not working hard enough.

"Are you winding me up, Mr Ho?"

"No, Stephanie."

Stephanie held his gaze for a moment, then she shrugged.

"Ten minutes," she said into the phone, and put it down. Mr Ho began to lecture her.

"Your trouble, Stephanie, is you lazy. Think too much about boys and not enough about work. Tell me, you learn speak French at school?"

Stephanie came into the kitchen and leant on the counter as Mr Ho rinsed the black wok under the tap. Steam hissed up.

"Learn French?" said Stephanie. "Why should I when I can run Transl-8 or Lebab or Syntactix? Who learns languages nowadays?"

The wok was already back on the flame.

"Wise people. Look at me. I come to England, I learn to speak the language properly. I don't use computer or headset every day. I want to be treated as Englishman, not foreigner."

"I don't see what that's got to do with the ants."

"How you think they speak English?" Mr Ho dropped a handful of onions and pork into the wok. "You think ants learn for themselves?"

"No. But I didn't think ants could use computers."

"Of course they can't. But software is context sensitive. If you learned languages, you would understand. Words aren't enough, need context too. Like when you say to me 'nice shirt, Mr Ho', the other day. You being sarcastic girl, think I not realize."

"No I wasn't..."

"You think I not clever because I speak with accent. But translating software know about accent and everything else. Context is everything. And context get bigger and bigger. Not just tone of voice, but facial expression, set of body, the whole environment. What called the frame. And so translating software need to do more than just listen to voice, but also to read body language and look at whole environment. Ants part of environment. Small wonder software start to translate movements straight into phone message."

"But why are ants ordering Chinese food?"

"Because my food very tasty. I good cook. Very popular takeaway. You think they want to eat Mr Mahmood's horrible curry? Rancid ghee and chilli powder."

"No! You know that's not what I mean! Why aren't they out hunting for leaves or whatever it is ants eat?"

"Why don't you go out and hunt or farm for food? Takeaway is easier. And tastier. And more efficient for ants. Nest is thriving. Is now 100 yards across!"

"What? How come I didn't know about that?"

"You lazy girl. Only think about boys. Ants nest on television, you not see?"

"No! But... where do they get the money from?"

"Ant nest is near computer assembly factory. Ants ideal for manipulating small parts. Cheaper than machinery too. Ants don't ask much money. Only get enough to buy tasty Chinese food."

"That's ridiculous!"

"No it's not. Is animals adapting to new ecosystem. You not listen to biology either? Too busy speaking to boyfriend."

Mr Ho dropped a handful of pineapple and pepper into the wok. He shook it in a hissing of steam and rattle of frying.

Stephanie bit her lip. "Are all the animals ordering takeaways, then?"

"No. Most animals lazy. Won't move with times. Like relatives who stay at home and not grasp opportunities of Western world."

Mr Ho picked up a ladle and, tossing the contents of the pan all the while, spooned three lots of red sauce over the ingredients.

"You're winding me up, right?" said Stephanie. "You're seriously telling me that because translation software looks



at the whole context of language, it has enabled communication between species, the results of which are that whole ecosystems are changing due to animals exploiting new opportunities. The result of this is that an ant's nest is now ordering takeaway food bought using money it earned by working in a microprocessor plant, this being a more viable means of existence?"

"Elegantly put. See? You clever girl when you want to be."

"So why does that mean I should learn French?"

"Ants not know language properly, not know wider context, so just respond to need. Get exploited. All that work in computer plant, and only get paid enough to buy takeaway."

The sweet and sour pork was ready. Mr Ho expertly divided the contents of the wok between two silver trays.

"All done," he said. "Feed hungry ants. Very good customers. Ah! Here come delivery girl!"

"Delivery girl?" said Stephanie. "What happened to Adam?"

"Too expensive. And lazy. Prefer spend time chatting to you than delivering tasty food. Delilah more reliable."

Stephanie looked down at Delilah as she trotted into the little kitchen, tail wagging. She was very good, eyes to the front, not stopping to sniff anything. Mr Ho bagged the order, putting the handle into her mouth.

"Good girl," he said as the dog trotted out of the shop. "Work very hard."

Stephanie watched her go.

The next day, she signed up for French lessons.

Two of Tony Ballantyne's most recent stories appear in *The Year's Best SF 13*, edited by David Hartwell and Kathryn Cramer. His next novel, *Twisted Metal*, is published in May by Macmillan.